

CC The gene relating to drug tolerance can be modified e.g. with the
 CC transporter inhibitors, screened compounds, antibodies and antisense
 CC nucleotides. The transporter is capable of selectively transporting
 CC indocarbazole compounds extracellularly. The present sequence represents
 CC the amino acid sequence of human ABCG2 mutant 482T

XX Sequence 655 AA:

Query Match 100.0%; Score 3352; DB 5; Length 655;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGEGAVLSFHNICYRVKLSGFLPCRKPYE 60
 DB 1 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGEGAVLSFHNICYRVKLSGFLPCRKPYE 60
 QY 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILNGARPPANFKCN 120
 DB 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILNGARPPANFKCN 120
 QY 121 SGYVODDVVMGTLTVRENIQFSALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 180
 DB 121 SGYVODDVVMGTLTVRENIQFSALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 180
 QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSGKRTIIF 240
 DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSGKRTIIF 240
 QY 241 SIHQPRYSIFKLPDSLTLASGRLMFHPGAEALGYESAGYHCEAVNNPADFFLDING 300
 DB 241 SIHQPRYSIFKLPDSLTLASGRLMFHPGAEALGYESAGYHCEAVNNPADFFLDING 300
 QY 301 DSTAVNALNREDEPKATEIIEPSKODKPLIEKLAIEIYNVSFFYKETAELHQLSGEKKKK 360
 DB 301 DSTAVNALNREDEPKATEIIEPSKODKPLIEKLAIEIYNVSFFYKETAELHQLSGEKKKK 360
 QY 361 ITVFEKISYTTSCFCHOLRWVSKSPKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLNKDS 420
 DB 361 ITVFEKISYTTSCFCHOLRWVSKSPKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLNKDS 420
 QY 421 TGIQNRAGVIFELTTCNCESSVAVELFVVEKKLFHEIYISGYRRVSYFLGKLSDLP 480
 DB 421 TGIQNRAGVIFELTTCNCESSVAVELFVVEKKLFHEIYISGYRRVSYFLGKLSDLP 480
 QY 481 MTMLPSIIFTCIYFVFMGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVVATLL 540
 DB 481 MTMLPSIIFTCIYFVFMGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVVATLL 540
 QY 541 MTICFVFMIFSGILVNLTTIASWLSWLYFSIPRYGFTALQHNELGQFCGLNATGN 600
 DB 541 MTICFVFMIFSGILVNLTTIASWLSWLYFSIPRYGFTALQHNELGQFCGLNATGN 600
 QY 601 NPCNVATCTGEEVTLVVGIDLSFWMGLMKHVALACMIVITLTAYLKLFLKRTYS 655
 DB 601 NPCNVATCTGEEVTLVVGIDLSFWMGLMKHVALACMIVITLTAYLKLFLKRTYS 655

RESULT 2

AA15221 standard; protein; 663 AA.

AA15221;

09-NOV-1999 (first entry)

Breast Cancer Resistance Protein (BCRP).

breast cancer; drug resistance; ATP-binding cassette; ABC;

xenobiotic transporter; chemotherapy; mitochondrion; doxorubicin;

breast cancer resistance protein; BCRP.

Homo sapiens.

FH Key Location/Qualifiers
 FT Domain 87..95
 FT Domain /note="Walker A motif"
 FT Domain 221..236
 FT Domain /note="Phosphopantetheine site"
 FT Modified-site 345..347
 FT Modified-site /note="Glycosylation site on N"
 FT Region 405..422
 FT Region /label=TM1
 FT Modified-site 425..427
 FT Modified-site /note="Transmembrane region"
 FT Region 546..563
 FT Region /label=TM2
 FT Modified-site 564..566
 FT Modified-site /note="Glycosylation site on N"
 FT Modified-site 604..606
 FT Modified-site /note="Glycosylation site on N"
 FT Region 638..655
 FT Region /label=TM3

W09940110-A1.

12-AUG-1999.

05-FEB-1999; 99WO-US002577.

05-FEB-1998; 98US-0073763P.

(UYMA-) UNIV MARYLAND BALTIMORE.

Ross DD, Doyle LA, Abruzzo L;

WPI: 1999-494273/41.

DR N-PSDB; AA206360.

PT New breast cancer resistance protein useful for production of antibodies
 PT to inhibit resistance activity for enhancing chemotherapy treatment.

PS Claim 4; Fig 2a; 80pp; English.

XX The Breast Cancer Resistance Protein (BCRP) is an ATP-binding cassette
 CC (ABC) transporter protein. It has a molecular mass of approximately 72.3
 CC kilodaltons (kD) exclusive of any glycosylation. Expression of BCRP in
 CC drug sensitive human cancer cells confers resistance to mitoxantrone,
 CC doxorubicin, and daunorubicin, and reduces daunorubicin accumulation in
 CC the cloned transfected cells. The protein is useful for producing
 CC antibodies and antisense probes, which can be used to inhibit the
 CC activity of BCRP, therefore enhancing a cancer patient's chemotherapy
 CC treatment. The antibodies and probes overcomes the problems of breast
 CC cancer resistance proteins to make chemotherapy treatment more effective

SQ Sequence 663 AA:

Query Match 100.0%; Score 3352; DB 2; Length 663;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGEGAVLSFHNICYRVKLSGFLPCRKPYE 60
 DB 9 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGEGAVLSFHNICYRVKLSGFLPCRKPYE 68
 QY 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILNGARPPANFKCN 120
 DB 69 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILNGARPPANFKCN 128
 QY 121 SGYVODDVVMGTLTVRENIQFSALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 180
 DB 129 SGYVODDVVMGTLTVRENIQFSALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 188
 QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSGKRTIIF 240
 DB 189 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSGKRTIIF 248

QY 241 SIHOPRYSIFKLPDSLTLLASGRLMFHGPAOEAALGFESAGYHCEAYNNPADFLDIING 300
DB 249 SIHOPRYSIFKLPDSLTLLASGRLMFHGPAOEAALGFESAGYHCEAYNNPADFLDIING 308
QY 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKXXX 360
DB 309 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKXXX 368
QY 361 ITVPEEISYTSFCHQLRWVSRSFKNLGNPOASIAOIIVTVVLGLVIGAIYFGKXDS 420
DB 369 ITVPEEISYTSFCHQLRWVSRSFKNLGNPOASIAOIIVTVVLGLVIGAIYFGKXDS 428
QY 421 TGIORAGVLPFLTNOCSSVSAVELFVVEKKLFHEHISGYRVSFELKLSDDL 480
DB 429 TGIORAGVLPFLTNOCSSVSAVELFVVEKKLFHEHISGYRVSFELKLSDDL 488
QY 481 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540
DB 489 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 548
QY 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPGLNATGN 600
DB 549 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPGLNATGN 608
QY 601 NPCNVAITCTGEBYLVKOGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 655
DB 609 NPCNVAITCTGEBYLVKOGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 663

RESULT 3
AAB60104
ID AAB60104 standard; protein; 655 AA.

AC AAB60104;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human transport protein TPPT-24.
XX
KW Human; transport protein; TPPT; transport disorder; metabolic disorder;
KW neurological disorder; cardiovascular disorder; reproductive disorder;
KW immune disorder; cancer.
OS Homo sapiens.
PN WO200078953-A2.
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US016668.
XX
PR 17-JUN-1999; 99US-0139923P.
PR 10-AUG-1999; 99US-0148177P.
PR 18-AUG-1999; 99US-0149357P.
PR 28-OCT-1999; 99US-0162287P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N,
PI Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;
XX
DR WPT; 2001-041424/05.
DR N-PSDB; AAF27724.
XX
PT Isolated polypeptide with a human transport protein sequence is useful
PT for the diagnosis, prevention and treatment of disorders associated with
PT the immune, reproductive and cardiovascular systems.
XX
PS Claim 2; Page 126-127; 165pp; English.
XX
CC The present invention provides the protein and coding sequences for 43
CC novel human transport proteins (designated TPPTs). These can be used in

CC the diagnosis and treatment of transport, metabolic, neurological,
CC reproductive, cardiovascular and immune disorders, and cell proliferative
CC disorders such as cancer

XX SQ Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 4; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTEGVLSFHNICRYKLGKGLPCPKPYE 60
DB 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTEGVLSFHNICRYKLGKGLPCPKPYE 60
QY 61 KEILSININGIMKPGNALITPTGGKSSLDVLAARKPSGLSGVLINGARPPANFKCN 120
DB 61 KEILSININGIMKPGNALITPTGGKSSLDVLAARKPSGLSGVLINGARPPANFKCN 120
QY 121 SGYVQDDVNGTTLVRENLOFSALRLATTNHEKNERINRVIOELGLDKVADSKYGT 180
DB 121 SGYVQDDVNGTTLVRENLOFSALRLATTNHEKNERINRVIOELGLDKVADSKYGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLLKRMSKQRTIIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLLKRMSKQRTIIF 240
QY 241 SIHOPRYSIFKLPDSLTLLASGRLMFHGPAOEAALGFESAGYHCEAYNNPADFLDIING 300
DB 241 SIHOPRYSIFKLPDSLTLLASGRLMFHGPAOEAALGFESAGYHCEAYNNPADFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKXXX 360
DB 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKXXX 360
QY 361 ITVPEEISYTSFCHQLRWVSRSFKNLGNPOASIAOIIVTVVLGLVIGAIYFGKXDS 420
DB 361 ITVPEEISYTSFCHQLRWVSRSFKNLGNPOASIAOIIVTVVLGLVIGAIYFGKXDS 420
QY 421 TGIORAGVLPFLTNOCSSVSAVELFVVEKKLFHEHISGYRVSFELKLSDDL 480
DB 421 TGIORAGVLPFLTNOCSSVSAVELFVVEKKLFHEHISGYRVSFELKLSDDL 480
QY 481 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540
DB 481 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540
QY 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPGLNATGN 600
DB 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPGLNATGN 600
QY 601 NPCNVAITCTGEBYLVKOGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 655
DB 601 NPCNVAITCTGEBYLVKOGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 655
RESULT 4
AA014781
ID AA014781 standard; protein; 655 AA.
XX
AC AA014781;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human BCRP protein.
XX
KW Human; BCRP protein; membrane penetrating region; cancer.
XX
OS Homo sapiens.
XX
PN JP2002065277-A.
PD 05-MAR-2002.
XX

PF 31-AUG-2000; 2000JP-00263742.
 XX
 PR 31-AUG-2000; 2000JP-00263742.
 XX
 PA (GANK-) ZH GAN KENKYUKAI.
 XX
 DR WPI; 2002-324198/36.
 DR N-PSDB; AAL42412.
 XX
 PT Mutant BCRP protein useful for treatment of cancer.
 XX
 PS Claim 13; Page 7-8; 15pp; Japanese.
 XX
 CC The invention comprises a mutant human BCRP protein, having a deletion,
 CC replacement or addition of at least one amino acid in the fifth membrane
 CC penetrating region of the wild-type BCRP protein. The mutant BCRP protein
 CC can be used for the treatment of cancer. The present amino acid sequence
 CC represents a human BCRP protein
 XX
 XQ Sequence 655 AA;

```

Db 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQSGGEKXXX 360
Qy 361 ITVFEISYTSFCHQLRWMSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420
Db 361 ITVFEISYTSFCHQLRWMSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420
Qy 421 TGIQRAGVLPFLTNOCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Db 421 TGIQRAGVLPFLTNOCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Qy 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540
Db 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540
Qy 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGNFCPLNATGN 600
Db 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGNFCPLNATGN 600
Qy 601 NPCNVAATCTGGEYLVKQIGIDLSPMGLMKHVAALACMIVIFLTITAYLKLFKKYS 655
Db 601 NPCNVAATCTGGEYLVKQIGIDLSPMGLMKHVAALACMIVIFLTITAYLKLFKKYS 655

RESULT 6
ABR58077
ID ABR58077 standard; protein; 655 AA.
XX
AC ABR58077;
XX
DT 15-OCT-2003 (first entry)
XX
DE Human ABCG2 protein.
XX
KW ABCG2; antidiabetic; cell therapy; diabetes mellitus;
XX pancreatic stem cell; islets of Langerhans; insulin.
XX
OS Homo sapiens.
XX
PN WO2003026584-A2.
XX
PD 03-APR-2003.
XX
PE 26-SEP-2002; 2002MO-US030700.
XX
PR 26-SEP-2001; 2001US-00963875.
XX 11-APR-2002; 2002US-00120687.
XX 02-MAY-2002; 2002US-00136891.
XX
PA (GENO) GEN HOSPITAL CORP.
XX
PI Habener JF, Zulewski H, Thomas MK, Abraham EJ, Vallejo M;
PI Leech CA, Nolan AL, Lechner A;
XX
DR MPI; 2003-354625/33.
XX
DR N-PBDB; ACC80605.
XX
PT Treating a patient with diabetes mellitus by isolating a nestin- or ABCG2
PT -positive pancreatic stem cell from a pancreatic islet of a donor and
PT transferring the stem cell into the patient.
XX
PS Disclosure; Fig 18B; 107pp; English.
XX
CC The invention relates to a method of treating a patient with diabetes
CC mellitus by isolating a nestin- or ABCG2-positive pancreatic stem cell
CC from a pancreatic islet of a donor, and transferring the stem cell into
CC the patient whereby the stem cell differentiates into an insulin-
CC producing cell. Alternatively, the nestin- or ABCG2-positive stem is
CC induced into a pancreatic progenitor cell prior to isolation and
CC transfer. This sequence corresponds to the human ABCG2 protein and the
CC encoding gene is detected in the method of the invention. The method is
CC useful for preparing a pharmaceutical composition for treating diabetes
CC mellitus. The stem cells can be further characterised for correct gene

```

```

CC expression using the primers and probes ACC80607-ACC80671
XX
SQ Sequence 655 AA;
XX
Query Match 99.8%; Score 3346; DB 6; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSSNVEVFIPVSOQNTNGFPATASNDLKAFEGAVLSFHNICVRVKLSGFLPCPKPYE 60
Db 1 MSSNVEVFIPVSOQNTNGFPATASNDLKAFEGAVLSFHNICVRVKLSGFLPCPKPYE 60
Qy 61 KEILSININGIMKPGNALILPTGCGKSSLDVLAARKDPSGSLGVLLINGARPNPFKN 120
Db 61 KEILSININGIMKPGNALILPTGCGKSSLDVLAARKDPSGSLGVLLINGARPNPFKN 120
Qy 121 SGYVQDDVVGTLTVRENLOPSAALRLATTTNHEKNERIRNVIQELGLDKVADSKVGT 180
Db 121 SGYVQDDVVGTLTVRENLOPSAALRLATTTNHEKNERIRNVIQELGLDKVADSKVGT 180
Qy 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLRMSKOGRTTF 240
Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLRMSKOGRTTF 240
Qy 241 SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGFESAGYHCAVNNPADFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGFESAGYHCAVNNPADFLDIING 300
Qy 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQSGGEKXXX 360
Db 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQSGGEKXXX 360
Qy 361 ITVFEISYTSFCHQLRWMSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420
Db 361 ITVFEISYTSFCHQLRWMSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420
Qy 421 TGIQRAGVLPFLTNOCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Db 421 TGIQRAGVLPFLTNOCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Qy 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540
Db 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540
Qy 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGNFCPLNATGN 600
Db 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGNFCPLNATGN 600
Qy 601 NPCNVAATCTGGEYLVKQIGIDLSPMGLMKHVAALACMIVIFLTITAYLKLFKKYS 655
Db 601 NPCNVAATCTGGEYLVKQIGIDLSPMGLMKHVAALACMIVIFLTITAYLKLFKKYS 655

RESULT 7
ADA10917
ID ADA10917 standard; protein; 655 AA.
XX
AC ADA10917;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human CDNA differentially expressed in colon cancer #23 product.
XX
KW differential expression; colon cancer; cancer; human.
XX
OS Homo sapiens.
XX
PN US2002160382-A1.
XX
PD 31-OCT-2002.
XX
PF 11-OCT-2001; 2001US-00981353.
XX

```

```
PR 11-OCT-2000; 2000US-0239841P.
XX (LASE/) LASEK A W.
PA (JONE/) JONES D A.
XX
XX Lasek AW, Jones DA;
XX WPI; 2003-265756/26.
XX N-PSDB; ADA10916.
XX
XX New combination comprising cDNAs that are differentially expressed in
PT colon disorder, useful for diagnosing, treating, staging or monitoring
XX treatment for colon cancers.
XX
XX Example 14; SEQ ID NO 35; 231bp; English.
XX
XX The invention relates to a combination comprising cDNAs that are
XX differentially expressed in colon disorder. The methods and compositions
XX of the present invention are useful for diagnosing, treating, staging or
XX monitoring treatment for colon cancer. They are also useful in high
XX throughput methods for using cDNAs to detect differential expression of
XX nucleic acids in a sample, screening molecules or compounds to identify a
XX ligand which specifically binds a cDNA and using a protein to screen
XX molecules or compounds to identify at least one ligand which specifically
XX binds the protein. The present sequence represents the amino acid
XX sequence of a human cDNA differentially expressed in colon cancer
XX protein.
XX
XX Sequence 655 AA:
SQ
Query Match 99.8%; Score 3346; DB 6; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYRLKSGFLPCRKYVE 60
DB 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYRLKSGFLPCRKYVE 60
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
QY 121 SGYVODDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKYADSKVGT 180
DB 121 SGYVODDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKYADSKVGT 180
QY 121 SGYVODDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKYADSKVGT 180
DB 121 SGYVODDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKYADSKVGT 180
QY 181 QFIRGVSGBGRKRTSIGMELITDPSILFDEPTGDSSTANVLLLLKMSQGRITIF 240
DB 181 QFIRGVSGBGRKRTSIGMELITDPSILFDEPTGDSSTANVLLLLKMSQGRITIF 240
QY 241 SIHQPRYSIFKLPDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFLDING 300
DB 241 SIHQPRYSIFKLPDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFLDING 300
QY 241 SIHQPRYSIFKLPDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFLDING 300
DB 241 SIHQPRYSIFKLPDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFLDING 300
QY 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIYVNSFYKETAELHQLSGEKKK 360
DB 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIYVNSFYKETAELHQLSGEKKK 360
QY 361 ITVFKESITSTSCHOIRAWSKRSPKULGNPOASTAQIIVTVVLGIVGAIYFGLKND 420
DB 361 ITVFKESITSTSCHOIRAWSKRSPKULGNPOASTAQIIVTVVLGIVGAIYFGLKND 420
QY 421 TGIQNBAGVLFLLTNOCFSSVAVELFVEKKLFIEHYISGYRVSSYFLGKLSDLPL 480
DB 421 TGIQNBAGVLFLLTNOCFSSVAVELFVEKKLFIEHYISGYRVSSYFLGKLSDLPL 480
QY 481 MTMLPSIIFTCTIYFMLGLPKADAFVMMFTLMVAVSASSMALAIACQSVSVATLL 540
DB 481 MTMLPSIIFTCTIYFMLGLPKADAFVMMFTLMVAVSASSMALAIACQSVSVATLL 540
QY 541 MTICFVFMMLFSGLVNLTITIASWLSWLOFSTIPRGFTALQNEPFGQFCGLNATGN 600
DB 541 MTICFVFMMLFSGLVNLTITIASWLSWLOFSTIPRGFTALQNEPFGQFCGLNATGN 600
```

```
QY 601 NPCNATCTGEBEYLKQIDLSFWMGLMKHVALACMIVIFLTIAYLKLFLKXYS 655
DB 601 NPCNATCTGEBEYLKQIDLSFWMGLMKHVALACMIVIFLTIAYLKLFLKXYS 655
XX
XX RESULT 8
XX ADC54182
XX ID ADC54182 standard; protein; 655 AA.
XX
XX ADC54182;
XX
XX 18-DEC-2003 (first entry)
XX
XX Human breast cancer resistance protein (BCRP) amino acid sequence.
XX
XX Cancer cell; anti-cancer agent; steroid hormone; oestrogenic effect;
XX BCRP; breast cancer resistance protein; cytosolic; camptothecin;
XX mitoxantrone; 7-hydroxy staurosporine; adriamycin; cancer chemotherapy;
XX human.
XX
XX Homo sapiens.
XX
XX JP2003063989-A.
XX
XX 05-MAR-2003.
XX
XX 23-AUG-2001; 2001JP-00252953.
XX
XX 23-AUG-2001; 2001JP-00252953.
XX
XX (GANK-) ZH GAN KENKYUKAI.
XX
XX WPI; 2003-735321/70.
XX
XX N-PSDB; ADC54181.
XX
XX Agent that overcomes resistance of cancer cell against anti-cancer agent,
XX comprises a steroid hormone, or a compound which exhibits antagonistic
XX activity against the hormone, with the cancer cell expressing BCRP gene.
XX
XX Example 1; SEQ ID NO 4; 15bp; Japanese.
XX
XX This invention relates to a novel agent which overcomes resistance of a
XX cancer cell against an anti-cancer agent (AA), comprising as an active
XX ingredient a steroid hormone, a compound having oestrogenic effect, or a
XX compound which exhibits antagonistic activity against the hormone, where
XX the cancer cell expresses the BCRP (breast cancer resistance protein)
XX gene. The agent of the invention may have cytostatic activity. The
XX invention is useful for overcoming resistance of a cancer against an anti
XX -cancer agent such as camptothecin, mitoxantrone, 7-hydroxy
XX staurosporine and adriamycin. The therapeutic effective anti-cancer agent
XX is recovered, due to the use of the agent of the invention. Also the
XX dosages of anti-cancer agent can be maintained easily, and adverse
XX effects of cancer chemotherapy can be suppressed. The present sequence is
XX that of the human BCRP protein which was used to develop the novel agent
XX of the invention.
XX
XX Sequence 655 AA:
SQ
Query Match 99.8%; Score 3346; DB 7; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYRLKSGFLPCRKYVE 60
DB 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYRLKSGFLPCRKYVE 60
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGSLVDLINGAPRPAFKN 120
QY 121 SGYVODDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKYADSKVGT 180
DB 121 SGYVODDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKYADSKVGT 180
```

```

Db      121. SGYVQDDVWGTLTVRENLQFSALRLATTTWTHKEKNERINRVOELGLDKVADSKVGT 180
Qy      181. QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGIDSSSTANAVALLLKRMKSGKRTIIF 240
Db      181. QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGIDSSSTANAVALLLKRMKSGKRTIIF 240
Qy      241. SIHOPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Db      241. SIHOPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Qy      301. DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETVAELHQSSEKSKK 360
Db      301. DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETVAELHQSSEKSKK 360
Qy      361. ITVFKETSYTTSPCHQLRWKSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Db      361. ITVFKETSYTTSPCHQLRWKSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Qy      421. TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Db      421. TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Qy      481. MTMLPSIIFTCIYFEMGLKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Db      481. MTMLPSIIFTCIYFEMGLKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Qy      541. MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Db      541. MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Qy      601. NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLLKYS 655
Db      601. NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLLKYS 655

```

RESULT 9

ADG38394
ID ADG38394 standard; protein; 655 AA.

ADG38394;

26-FEB-2004 (first entry)

Human wild-type BCRP.

Anticancer agent; polymorphism; human; BCRP; cancer cell.

Homo sapiens.

JP2003199585-A.

15-JUL-2003.

21-MAY-2002; 2002JP-00145926.

24-OCT-2001; 2001JP-00325883.

(GANK-) ZH GAN KENKUKAI.

WPI; 2003-819597/77.

N-PSDB; ADG38395.

Evaluating sensitivity of test cell to anticancer agent involves

identifying gene polymorphism of BCRP.

Example 1; SEQ ID NO 7; 18bp; Japanese.

The present invention relates to a method for evaluating the sensitivity of a cell to an anticancer agent. The method involves identifying a gene polymorphism in the human BCRP gene (the polymorphism is undefined in the specification). The gene polymorphisms encode variant BCRP polypeptides designated as Q141K, V14M and Q126STOP. Identifying the gene polymorphism of BCRP of a test cell is useful for evaluating the expression grade of

CC the side effect at the time of administering an anticancer agent to the
CC test cell and evaluating the resistance of the test cell to the
CC anticancer agent. BCRP protein is useful in conveying an anticancer agent
CC to cancer cell. The method is efficient in identifying a safer anticancer
CC agent for treatment. The present sequence represents wild-type BCRP.

XX Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 7; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1. MSSNVEVPIPSQGTNGFPATASNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPVE 60
Db      1. MSSNVEVPIPSQGTNGFPATASNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPVE 60
Qy      61. KEILSNINGIMKPGINALIGPTGGKSSLDVLAARKDPSGLSDVLLNGAPRPANFKCN 120
Db      61. KEILSNINGIMKPGINALIGPTGGKSSLDVLAARKDPSGLSDVLLNGAPRPANFKCN 120
Qy      121. SGYVQDDVWGTLTVRENLQFSALRLATTTWTHKEKNERINRVOELGLDKVADSKVGT 180
Db      121. SGYVQDDVWGTLTVRENLQFSALRLATTTWTHKEKNERINRVOELGLDKVADSKVGT 180
Qy      181. QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGIDSSSTANAVALLLKRMKSGKRTIIF 240
Db      181. QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGIDSSSTANAVALLLKRMKSGKRTIIF 240
Qy      241. SIHOPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Db      241. SIHOPRYSIFKLFDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Qy      301. DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETVAELHQSSEKSKK 360
Db      301. DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETVAELHQSSEKSKK 360
Qy      361. ITVFKETSYTTSPCHQLRWKSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Db      361. ITVFKETSYTTSPCHQLRWKSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Qy      421. TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Db      421. TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Qy      481. MTMLPSIIFTCIYFEMGLKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Db      481. MTMLPSIIFTCIYFEMGLKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Qy      541. MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Db      541. MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Qy      601. NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLLKYS 655
Db      601. NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLLKYS 655

```

RESULT 10

ADIS7316
ID ADIS7316 standard; protein; 655 AA.

ADIS7316;

22-APR-2004 (first entry)

ATP-binding cassette transporter ABCG2 D590Y mutant.

drug transport capability; polymorphism; ABCG2; polymorphic mutation;

drug sensitivity; anti-cancer drug; cancer therapy;

cancer cell detection; indolocarbozole compound; human;

ABC transporter superfamily;

ATP-binding cassette transporter superfamily; mutant; mutenin.

OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 590 /note= "Wild type Asp substituted by Tyr"
 XX
 PN WO2003107249-A1.
 XX
 PD 24-DEC-2003.
 XX
 PF 13-JUN-2003; 2003WO-JP007534.
 XX
 PR 17-JUN-2002; 2002JP-00175806.
 XX
 PA (BANY) BANYU PHARM CO LTD.
 XX
 PI Kotani H, Mizunara S;
 DR WPI; 2004-156349/15.
 XX
 PT Predicting drug transport capability of mammalian cell by collecting
 sample from mammal, determining polymorphism of nucleotide sequence of
 ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.
 XX
 PS Example 1; Page; 76pp; English.
 XX
 CC The invention describes a method of predicting a drug transport
 capability of a mammalian cell involving collecting a sample from a
 mammal, determining a polymorphism of the nucleotide sequence of ABCG2
 gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide.
 CC The method is useful for predicting drug transport capability of a
 mammalian cell. Polypeptides comprising single nucleotide
 CC polymorphisms or polypeptides comprising polymorphic mutations of the
 ABCG2 protein are useful as diagnostic agent for diagnosing drug
 CC sensitivity which involves analyzing a biological sample from a subject
 CC and determining the presence or absence of the polymorphisms or
 CC polypeptides, where the subject having the polymorphisms and/or the
 CC polypeptide is suggested to be sensitive to the indolocarbazole compound.
 CC A transformed cell comprising an ABCG2 protein mutant is useful for
 CC measuring drug transport capability. By predicting drug transport
 CC capability of a mammalian cell, sensitivity of a patient to various drugs
 CC such as anti-cancer drugs can be diagnosed and an indicator for the
 CC therapy can be obtained. As a result of selecting an anti-cancer drug in
 CC cancer therapy and, particularly, detecting a cancer cell(s) which is
 CC highly sensitive to indolocarbazole compounds, it is now possible to
 CC selectively apply the compounds for the therapy. In addition, the optimum
 CC dose of the indolocarbazole compounds in the cancer therapy is found and,
 CC at the same time, side effect of the compounds is reduced whereby a
 CC highly effective method of using the indolocarbazole compounds is
 CC provided. This is the amino acid sequence of a human ABC transporter
 CC superfamily (ATP-binding cassette transporter superfamily) protein ABCG2
 CC mutant. Note: This information does not appear in the specification but has
 CC been created using information given in the claims of the invention.
 XX
 SQ Sequence 655 AA;
 XX

Query Match 99.8%; Score 3346; DB 8; Length 655;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEFPIVSGNTNGFPATASNDLKAFTEGAVLSFNNICRYVLKSGFLPCRPE 60
 DB 1 MSSSNVEFPIVSGNTNGFPATASNDLKAFTEGAVLSFNNICRYVLKSGFLPCRPE 60
 QY 61 KEILSNINGIMKGLNAILPTGGGKSLDVLAAKDPGSLGSDVILNGAPPPANFKCN 120
 DB 61 KEILSNINGIMKGLNAILPTGGGKSLDVLAAKDPGSLGSDVILNGAPPPANFKCN 120
 QY 121 SGVVDVDMGTLTVRENIQFSAALRLATMTMHEKNERINRYIGELGIDKYADSVGT 180
 DB 121 SGVVDVDMGTLTVRENIQFSAALRLATMTMHEKNERINRYIGELGIDKYADSVGT 180

QY 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSPANAVLLLLKMSKQGRITIF 240
 DB 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSPANAVLLLLKMSKQGRITIF 240
 QY 241 SIHQPRYSIFKLPDSLTLLASGRLMFRGPAQELAGFESAGHYCEAVNNPADFLDIING 300
 DB 241 SIHQPRYSIFKLPDSLTLLASGRLMFRGPAQELAGFESAGHYCEAVNNPADFLDIING 300
 QY 301 DSTAVALLNREDEKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
 DB 301 DSTAVALLNREDEKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
 QY 361 ITVFKESITSTSPCHQLRWKSKSPKULLGNPQASIAQIIVTVVLGIVGAIYFGKND 420
 DB 361 ITVFKESITSTSPCHQLRWKSKSPKULLGNPQASIAQIIVTVVLGIVGAIYFGKND 420
 QY 421 TGIQNRAGVLFLLTQOCSSVSVAVELFVEKKLFHEYSIGYRVSSYFLGLSDLLP 480
 DB 421 TGIQNRAGVLFLLTQOCSSVSVAVELFVEKKLFHEYSIGYRVSSYFLGLSDLLP 480
 QY 481 MTMLPSIIFTCIYFMGLGKPKADAFVMMFTLMVAYASSMALAIAGQSVSVATLL 540
 DB 481 MTMLPSIIFTCIYFMGLGKPKADAFVMMFTLMVAYASSMALAIAGQSVSVATLL 540
 QY 541 MTICFVFMIFSGILVNLTTIASWLSMLOYFSIPRYGFTALQNEFLQNFCEGLNATGN 600
 DB 541 MTICFVFMIFSGILVNLTTIASWLSMLOYFSIPRYGFTALQNEFLQNFCEGLNATGN 600
 QY 601 NPGNYATCTGEEYLVAGIDLSPPWGLMKHVALACMIVIFLTAYLKLFLKRTS 655
 DB 601 NPGNYATCTGEEYLVAGIDLSPPWGLMKHVALACMIVIFLTAYLKLFLKRTS 655

RESULT 11
 ADI57315
 ID ADI57315 standard; protein; 655 AA.
 AC ADI57315;
 XX
 DT 22-APR-2004 (first entry)
 XX
 DE ATP-binding cassette transporter ABCG2 R482T mutant.
 XX
 KW drug transport capability; polymorphism; ABCG2; polymorphic mutation;
 KW drug sensitivity; anti-cancer drug; cancer therapy;
 KW cancer cell detection; indolocarbazole compound; human;
 KW ABC transporter superfamily;
 KW ATP-binding cassette transporter superfamily; mutant; muten.
 XX
 OS Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 482 /note= "Wild type Arg substituted by Thr"
 FT
 PN WO2003107249-A1.
 XX
 PD 24-DEC-2003.
 XX
 PF 13-JUN-2003; 2003WO-JP007534.
 XX
 PR 17-JUN-2002; 2002JP-00175806.
 XX
 PA (BANY) BANYU PHARM CO LTD.
 XX
 PI Kotani H, Mizunara S;
 DR WPI; 2004-156349/15.
 XX
 PT Predicting drug transport capability of mammalian cell by collecting
 sample from mammal, determining polymorphism of nucleotide sequence of
 ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.
 PT

Example 1; Page: 76pp; English.

The invention describes a method of predicting a drug transport capability of a mammalian cell involving collecting a sample from a mammal, determining a polymorphism of the nucleotide sequence of ABCG2 gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide. The method is useful for predicting drug transport capability of a mammalian cell. Polynucleotides comprising single nucleotide polymorphisms or polypeptides comprising polymorphic mutations of the ABCG2 protein are useful as diagnostic agent for diagnosing drug sensitivity which involves analyzing a biological sample from a subject and determining the presence or absence of the polynucleotides or polypeptides, where the subject having the polynucleotide and/or the polypeptide is suggested to be sensitive to the indolocarbazole compound. A transformed cell comprising an ABCG2 protein mutant is useful for measuring drug transport capability. By predicting drug transport capability of a mammalian cell, sensitivity of a patient to various drugs such as anti-cancer drugs can be diagnosed and an indicator for the therapy can be obtained. As a result of selecting an anti-cancer drug in cancer therapy and, particularly, detecting a cancer cell(s) which is highly sensitive to indolocarbazole compounds, it is now possible to selectively apply the compounds for the therapy. In addition, the optimum dose of the indolocarbazole compounds in the cancer therapy is found and, at the same time, side effect of the compounds is reduced whereby a highly effective method of using the indolocarbazole compounds is provided. This is the amino acid sequence of a human ABC transporter superfamily (ATP-binding cassette transporter superfamily) protein ABCG2 mutant. Note: This sequence does not appear in the specification but has been created using information given in the claims of the invention.

Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 8; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 MSSNVEVPIPIPSQNTNGFPATASNDLKAFTGAVLSFPHNICYVVKLSGSLPCKPVE 60
DB 1 MSSNVEVPIPIPSQNTNGFPATASNDLKAFTGAVLSFPHNICYVVKLSGSLPCKPVE 60
OY 61 KEILNININGIMKPGNALIGPTGGGKSSLDVLAARKDPSGSLVDLNGAPRPNFKCN 120
DB 61 KEILNININGIMKPGNALIGPTGGGKSSLDVLAARKDPSGSLVDLNGAPRPNFKCN 120
OY 121 SGYVVDVDMVMTLTRENLQFSALRLATTMTNHEKNERINRVIOELGLDQVADSKVGT 180
DB 121 SGYVVDVDMVMTLTRENLQFSALRLATTMTNHEKNERINRVIOELGLDQVADSKVGT 180
OY 181 QPIRVSNGGERKRTSIGMELITDPSILFDEPTTGIDSTANAVALLLKRMKSKQRTIIF 240
DB 181 QPIRVSNGGERKRTSIGMELITDPSILFDEPTTGIDSTANAVALLLKRMKSKQRTIIF 240
OY 241 SIHQRYSTFVKLFDSTITLASGRLMFHGPAQALGFESAGVHCENYNNPAPFDIING 300
DB 241 SIHQRYSTFVKLFDSTITLASGRLMFHGPAQALGFESAGVHCENYNNPAPFDIING 300
OY 301 DSTAVANLNREDFKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLQSGEKKKK 360
DB 301 DSTAVANLNREDFKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLQSGEKKKK 360
OY 361 ITVFEISYTTTSFCHQLRWVSKSRNLGNPOASIAQIIVVVGIVGAIYFGDKNDS 420
DB 361 ITVFEISYTTTSFCHQLRWVSKSRNLGNPOASIAQIIVVVGIVGAIYFGDKNDS 420
OY 421 TGIQRAGVLPFLTNQCPSSSAVELFVEKKLFHEIYSQYRRSSYFLKSLDILLP 480
DB 421 TGIQRAGVLPFLTNQCPSSSAVELFVEKKLFHEIYSQYRRSSYFLKSLDILLP 480
OY 481 MTMLPSIIFTCTIVYFMLGLKPRDAEFVMMFTLMVAVYSASSMALAIAAGQSVSVATLL 540
DB 481 MTMLPSIIFTCTIVYFMLGLKPRDAEFVMMFTLMVAVYSASSMALAIAAGQSVSVATLL 540

```

```

OY 541 MTICFVPMWISGGLLVNLTITIASWLSWLQYFSIPRYGFTALQHNELFQNCFCGINATCN 600
DB 541 MTICFVPMWISGGLLVNLTITIASWLSWLQYFSIPRYGFTALQHNELFQNCFCGINATCN 600
OY 601 NPGNATCTGSEYLYKOGIDLSFNGLMKNHVALACMIVIFLITAVLKLLFLKKYS 655
DB 601 NPGNATCTGSEYLYKOGIDLSFNGLMKNHVALACMIVIFLITAVLKLLFLKKYS 655

RESULT 12
AD157243
ID AD157243 standard; protein; 655 AA.
AC AD157243;
DT 22-APR-2004 (first entry)
DE Human ATP-binding cassette transporter ABCG2.
KW drug transport capability; polymorphism; ABCG2; polymorphic mutation;
KW cancer cell detection; indolocarbazole compound; cancer therapy;
KW ABC transporter superfamily;
KW ATP-binding cassette transporter superfamily.
OS Homo sapiens.
PN WO2003107249-A1.
PD 24-DEC-2003.
PF 13-JUN-2003; 2003WO-JP007534.
PR 17-JUN-2002; 2002JP-00175806.
PA (BANY ) BANYU PHARM CO LTD.
PI Kotani H, Mizutani S;
DR WPI, 2004-156349/15.
XX N-PSDB; AD157242.
PT Predicting drug transport capability of mammalian cell by collecting
PT sample from mammal, determining polymorphism of nucleotide sequence of
PT ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.
XX
XX Claim 16; SEQ ID NO 2; 76pp; English.
PS
CC The invention describes a method of predicting a drug transport
CC capability of a mammalian cell involving collecting a sample from a
CC mammal, determining a polymorphism of the nucleotide sequence of ABCG2
CC gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide.
CC The method is useful for predicting drug transport capability of a
CC mammalian cell. Polynucleotides comprising single nucleotide
CC polymorphisms or polypeptides comprising polymorphic mutations of the
CC ABCG2 protein are useful as diagnostic agent for diagnosing drug
CC sensitivity which involves analyzing a biological sample from a subject
CC and determining the presence or absence of the polynucleotides or
CC polypeptides, where the subject having the polynucleotide and/or the
CC polypeptide is suggested to be sensitive to the indolocarbazole compound.
CC A transformed cell comprising an ABCG2 protein mutant is useful for
CC measuring drug transport capability. By predicting drug transport
CC capability of a mammalian cell, sensitivity of a patient to various drugs
CC such as anti-cancer drugs can be diagnosed and an indicator for the
CC cancer therapy can be obtained. As a result of selecting an anti-cancer drug in
CC cancer therapy and, particularly, detecting a cancer cell(s) which is
CC highly sensitive to indolocarbazole compounds, it is now possible to
CC selectively apply the compounds for the therapy. In addition, the optimum
CC dose of the indolocarbazole compounds in the cancer therapy is found and,
CC at the same time, side effect of the compounds is reduced whereby a
CC highly effective method of using the indolocarbazole compounds is
CC provided. This is the amino acid sequence of human ABC transporter
CC superfamily (ATP-binding cassette transporter superfamily) protein ABCG2.

```

XX SQ Sequence 655 AA;
 Query Match 99.8%; Score 3346; DB 8; Length 655;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKYE 60
 |||||
 DB 1 MSSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKYE 60
 |||||

QY 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDLINGAPPANFKCN 120
 |||||
 DB 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDLINGAPPANFKCN 120
 |||||

QY 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180
 |||||
 DB 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180
 |||||

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
 |||||
 DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
 |||||

QY 241 SIHQPRYSIFKLFDSLTLLASGRIMFHPAQEALGYFESAGYHCEAYNNPADPFLDIING 300
 |||||
 DB 241 SIHQPRYSIFKLFDSLTLLASGRIMFHPAQEALGYFESAGYHCEAYNNPADPFLDIING 300
 |||||

QY 301 DSTAVALNREEDPKATEIIEPSKODKPLEKLEIYVNSFEYKETAELHQLSGGEKKK 360
 |||||
 DB 301 DSTAVALNREEDPKATEIIEPSKODKPLEKLEIYVNSFEYKETAELHQLSGGEKKK 360
 |||||

QY 361 ITVFKEISYTTSEFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKLNDS 420
 |||||
 DB 361 ITVFKEISYTTSEFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKLNDS 420
 |||||

QY 421 TGIQNRAGVLPFLTTNOCSSVASVBLFVVEKLFHEIYSGYRRVSYFLGKLSDDL 480
 |||||
 DB 421 TGIQNRAGVLPFLTTNOCSSVASVBLFVVEKLFHEIYSGYRRVSYFLGKLSDDL 480
 |||||

QY 481 MTMLPSIIFTCIYFMYLGLKPKADAFVMMFTLMMVAYSSSALAIAGQSVSVATLL 540
 |||||
 DB 481 MTMLPSIIFTCIYFMYLGLKPKADAFVMMFTLMMVAYSSSALAIAGQSVSVATLL 540
 |||||

QY 541 MTICFVPMIFSGILVNLTTIASWLSLQYFSIPRYGFTALQHNBPGLQNFCEGLNATGN 600
 |||||
 DB 541 MTICFVPMIFSGILVNLTTIASWLSLQYFSIPRYGFTALQHNBPGLQNFCEGLNATGN 600
 |||||

QY 601 NPCNVATCTGGEYLVKQIDLSPWGLMKHVALACMIVIFLTTAYLKLFLKXYS 655
 |||||
 DB 601 NPCNVATCTGGEYLVKQIDLSPWGLMKHVALACMIVIFLTTAYLKLFLKXYS 655
 |||||

RESULT 13
 ADI57311
 ID ADI57311 standard; protein; 655 AA.
 XX
 XX ADI57311;
 DT 22-APR-2004 (first entry)
 XX
 XX ATP-binding cassette transporter ABCG2 Q141K mutant.
 DE
 XX drug transport capability; polymorphism; ABCG2; polymorphic mutation;
 KM drug sensitivity; anti-cancer drug; cancer therapy;
 KM cancer cell detection; indolocarbazole compound; human;
 KM ABC transporter superfamily;
 XX ATP-binding cassette transporter superfamily; mutant; mutein.
 OS
 XX Homo sapiens.
 OS
 XX Synthetic.
 Key Location/Qualifiers
 FT Misc-difference 141

FT /note="Wild type Gln substituted by Lys"
 XX
 XX WO2003107249-A1.
 XX
 XX 24-DEC-2003.
 PD
 XX 13-JUN-2003; 2003WO-UP007534.
 PF
 XX 17-JUN-2002; 2002JP-00175806.
 PR
 XX (BANY) BANYU PHARM CO LTD.
 PA
 XX Kotani H, Mizunari S;
 PI
 XX WPI; 2004-156349/15.
 DR
 XX
 XX
 XX Predicting drug transport capability of mammalian cell by collecting
 PT sample from mammal, determining polymorphism of nucleotide sequence of
 PT ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.
 XX
 XX Claim 6; Page; 76pp; English.
 PS
 CC The invention describes a method of predicting a drug transport
 CC capability of a mammalian cell involving collecting a sample from a
 CC mammal, determining a polymorphism of the nucleotide sequence of ABCG2
 CC gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide.
 CC The method is useful for predicting drug transport capability of a
 CC mammalian cell. Polynucleotides comprising single nucleotide
 CC polymorphisms or polypeptides comprising polymorphic mutations of the
 CC ABCG2 protein are useful as diagnostic agent for diagnosing drug
 CC sensitivity which involves analyzing a biological sample from a subject
 CC and determining the presence or absence of the polynucleotides or
 CC polypeptides, where the subject having the polynucleotide and/or the
 CC polypeptide is suggested to be sensitive to the indolocarbazole compound.
 CC A transformed cell comprising an ABCG2 protein mutant is useful for
 CC measuring drug transport capability. By predicting drug transport
 CC capability of a mammalian cell, sensitivity of a patient to various drugs
 CC such as anti-cancer drugs can be diagnosed and an indicator for the
 CC therapy can be obtained. As a result of selecting an anti-cancer drug in
 CC cancer therapy and, particularly, detecting a cancer cell(s) which is
 CC highly sensitive to indolocarbazole compounds, it is now possible to
 CC selectively apply the compounds for the therapy. In addition, the optimum
 CC dose of the indolocarbazole compounds in the cancer therapy is found and,
 CC at the same time, side effect of the compounds is reduced whereby a
 CC highly effective method of using the indolocarbazole compounds is
 CC provided. This is the amino acid sequence of a human ABC transporter
 CC superfamily (ATP-binding cassette transporter superfamily) protein ABCG2
 CC mutant. Note: This sequence does not appear in the specification but has
 CC been created using information given in the claims of the invention.
 CC
 XX
 XX SQ Sequence 655 AA;
 Query Match 99.8%; Score 3346; DB 8; Length 655;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKYE 60
 |||||
 DB 1 MSSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKYE 60
 |||||

QY 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDLINGAPPANFKCN 120
 |||||
 DB 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDLINGAPPANFKCN 120
 |||||

QY 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180
 |||||
 DB 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180
 |||||

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
 |||||
 DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
 |||||

QY 241 SIHQPRYSIFKLFDSLTLLASGRIMFHPAQEALGYFESAGYHCEAYNNPADPFLDIING 300
 |||||

```

Db 241 SIHQPRYSIFKLPDSTLLASGRLMFHGPAQBALGYFESAGYHCEAYNNPADFLDIING 300
Qy 301 DSTAVALNREEDPKATEIIEPSKODKPLEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Db 301 DSTAVALNREEDPKATEIIEPSKODKPLEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Qy 361 ITVFKEISYTTSPCHQLRWMSKRSFKNLLGNPOASIAQIIVTVVLGIVGAIYFGLKXDS 420
Db 361 ITVFKEISYTTSPCHQLRWMSKRSFKNLLGNPOASIAQIIVTVVLGIVGAIYFGLKXDS 420
Qy 421 TGIQNRAGVLPFLTTNOCFSSVSAYELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Db 421 TGIQNRAGVLPFLTTNOCFSSVSAYELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Qy 481 MTMLPSIIIFTCIVYFMLGLKPKADAFVVMFTLMVAVASASMAIAAGOSVSVATLL 540
Db 481 MTMLPSIIIFTCIVYFMLGLKPKADAFVVMFTLMVAVASASMAIAAGOSVSVATLL 540
Qy 541 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONPCGLNATGN 600
Db 541 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONPCGLNATGN 600
Qy 601 NPCNVATCTGGEYLVKQIGIDLSPWGLMKXHVLAACMIVIFLTIAYLKLFLEKYS 655
Db 601 NPCNVATCTGGEYLVKQIGIDLSPWGLMKXHVLAACMIVIFLTIAYLKLFLEKYS 655

```

RESULT 14

ADK67372
ID ADK67372 standard; protein; 655 AA.

AC ADK67372;

DT 18-NOV-2004 (first entry)

DE Human wild-type ABCG2 (ATP-binding cassette gene) protein.

KM drug absorption; ABCG2; ATP-binding cassette gene; human; wild-type; chromosome 4q22.

XX Homo sapiens.

PN JP2004016042-A.

XX 22-JAN-2004.

PF 13-JUN-2002; 2002JP-00172759.

PR 13-JUN-2002; 2002JP-00172759.

PA (KOKU-) KOKURITSU IYAKUHIN SHOKUHIN EISEI KENKYU.

PA (IYAK-) IYAKUHIN FUKUSAYO HIGAI KYUSAI KENKYU SH.

DR WPI; 2004-113852/12.

DR N-PSDB; ADK67371.

XX Novel ABCG2 polynucleotide having a mutation at a specific position, associated with ABCG2 protein.

XX Claim 1; SEQ ID NO 2; 53pp; Japanese.

CC The invention relates to a novel polynucleotide having a mutation in the codon encoding a glutamine residue present at the 126 position of a 655 amino acid sequence. The polynucleotide of the invention may be useful for the estimation or diagnosis of a condition which is associated with abnormal drug absorption and in which the ABCG2 (ATP-binding cassette gene) protein is associated. The current sequence is that of the human wild-type ABCG2 protein of the invention which is encoded by DNA located at chromosome 4q22.

CC Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 8; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 MSSNVEVFIPIVSGNTNGPPTASNDLKAFTEGAVLSPFHNI CYRVKLSGFLPCKRVE 60
Db 1 MSSNVEVFIPIVSGNTNGPPTASNDLKAFTEGAVLSPFHNI CYRVKLSGFLPCKRVE 60
Qy 61 KEILSNINGIMKPGINALLPGTGGKSLDLVLAARKDPSGSLVINGARPRNPKCN 120
Db 61 KEILSNINGIMKPGINALLPGTGGKSLDLVLAARKDPSGSLVINGARPRNPKCN 120
Qy 121 SGYVQDDVVMGTLVRENLOFSALRLATTMTHEKNERIRNVOELGLDVAASKYGT 180
Db 121 SGYVQDDVVMGTLVRENLOFSALRLATTMTHEKNERIRNVOELGLDVAASKYGT 180
Qy 181 QPIRVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLKRMSSKQRTIIF 240
Db 181 QPIRVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLKRMSSKQRTIIF 240
Qy 241 SIHQPRYSIFKLPDSTLLASGRLMFHGPAQBALGYFESAGYHCEAYNNPADFLDIING 300
Db 241 SIHQPRYSIFKLPDSTLLASGRLMFHGPAQBALGYFESAGYHCEAYNNPADFLDIING 300
Qy 301 DSTAVALNREEDPKATEIIEPSKODKPLEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Db 301 DSTAVALNREEDPKATEIIEPSKODKPLEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Qy 361 ITVFKEISYTTSPCHQLRWMSKRSFKNLLGNPOASIAQIIVTVVLGIVGAIYFGLKXDS 420
Db 361 ITVFKEISYTTSPCHQLRWMSKRSFKNLLGNPOASIAQIIVTVVLGIVGAIYFGLKXDS 420
Qy 421 TGIQNRAGVLPFLTTNOCFSSVSAYELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Db 421 TGIQNRAGVLPFLTTNOCFSSVSAYELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Qy 481 MTMLPSIIIFTCIVYFMLGLKPKADAFVVMFTLMVAVASASMAIAAGOSVSVATLL 540
Db 481 MTMLPSIIIFTCIVYFMLGLKPKADAFVVMFTLMVAVASASMAIAAGOSVSVATLL 540
Qy 541 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONPCGLNATGN 600
Db 541 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONPCGLNATGN 600
Qy 601 NPCNVATCTGGEYLVKQIGIDLSPWGLMKXHVLAACMIVIFLTIAYLKLFLEKYS 655
Db 601 NPCNVATCTGGEYLVKQIGIDLSPWGLMKXHVLAACMIVIFLTIAYLKLFLEKYS 655

```

RESULT 15

ADIS7314
ID ADIS7314 standard; protein; 655 AA.

AC ADIS7314;

DT 22-APR-2004 (first entry)

DE ATP-binding cassette transporter ABCG2 R482G mutant.

XX drug transport capability; polymorphism; ABCG2; polymorphic mutation;

XX drug sensitivity; anti-cancer drug; cancer therapy;

XX cancer cell detection; indolocarbazole compound; human;

XX ABC transporter superfamily;

XX ATP-binding cassette transporter superfamily; mutant; mutein.

OS Homo sapiens.

XX Synthetic.

XX Key

FT Misc-difference 482

Location/Qualifiers
/note= "Wild type Arg substituted by Gly"

PN WO2003107249-A1.
 XX 24-DEC-2003.
 XX 13-JUN-2003; 2003WO-JP007534.
 PF 17-JUN-2002; 2002JP-00175806.
 XX (BANY) BANYU PHARM CO LTD.
 PA Kotani H, Mizuara S;
 PI WPI; 2004-156349/15.
 DR
 XX
 PT Predicting drug transport capability of mammalian cell by collecting
 sample from mammal, determining polymorphism of nucleotide sequence of
 PT ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.
 XX
 PS Example 1; Page: 76pp; English.
 XX
 CC The invention describes a method of predicting a drug transport
 CC capability of a mammalian cell involving collecting a sample from a
 CC mammal, determining a polymorphism of the nucleotide sequence of ABCG2
 CC gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide.
 CC The method is useful for predicting drug transport capability of a
 CC mammalian cell. Polynucleotides comprising single nucleotide
 CC polymorphisms or polynucleotides comprising polymorphic mutations of the
 CC ABCG2 protein are useful as diagnostic agent for diagnosing drug
 CC sensitivity which involves analyzing a biological sample from a subject
 CC and determining the presence or absence of the polynucleotides or
 CC polypeptides, where the subject having the polynucleotide and/or the
 CC polypeptide is suggested to be sensitive to the indolocarbazole compound.
 CC A transformed cell comprising an ABCG2 protein mutant is useful for
 CC measuring drug transport capability. By predicting drug transport
 CC capability of a mammalian cell, sensitivity of a patient to various drugs
 CC such as anti-cancer drugs can be diagnosed and an indicator for the
 CC therapy can be obtained. As a result of selecting an anti-cancer drug in
 CC cancer therapy and, particularly, detecting a cancer cell(s) which is
 CC highly sensitive to indolocarbazole compounds, it is now possible to
 CC selectively apply the compounds for the therapy. In addition, the optimum
 CC dose of the indolocarbazole compounds in the cancer therapy is found and,
 CC at the same time, side effect of the compounds is reduced whereby a
 CC highly effective method of using the indolocarbazole compounds is
 CC provided. This is the amino acid sequence of a human ABC transporter
 CC superfamily (ATP-binding cassette transporter superfamily) protein ABCG2
 CC mutant. Note: This sequence does not appear in the specification but has
 CC been created using information given in the claims of the invention.
 XX
 SO Sequence 655 AA.

Query Match 99.8%; Score 3345; DB 8; Length 655;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSSSNVEVFPVSGQNTNGPPTASNDLKAFTGCAVLSFNICRYVKLKSGLPCKRVE 60
 DB 1 MSSSNVEVFPVSGQNTNGPPTASNDLKAFTGCAVLSFNICRYVKLKSGLPCKRVE 60
 QY 61 KEILSNNGIMKPGLNAILPPTGGKSSLLDVLAAKDPGSLGSDVLIINGAPPANPKCN 120
 DB 61 KEILSNNGIMKPGLNAILPPTGGKSSLLDVLAAKDPGSLGSDVLIINGAPPANPKCN 120
 QY 121 SGYVODDVVVGTLVRENLFSAALRLATMTNHEKNERINRVTOELGLDKVADSKVGT 180
 DB 121 SGYVODDVVVGTLVRENLFSAALRLATMTNHEKNERINRVTOELGLDKVADSKVGT 180
 QY 181 QFIRGVSGGERKRTSIGMEILTPDSIIFLDEPTTGLDSTANAVLLILKMSKQGRITIF 240
 DB 181 QFIRGVSGGERKRTSIGMEILTPDSIIFLDEPTTGLDSTANAVLLILKMSKQGRITIF 240
 QY 241 SIHQPRYSIFKLPDSLITLILASGRLMFHPQAOBALGYFESAGYHCEAYNNPADPFLDITNG 300
 DB 241 SIHQPRYSIFKLPDSLITLILASGRLMFHPQAOBALGYFESAGYHCEAYNNPADPFLDITNG 300

QY 301 DSTAVALNREDEPKATEIIEPSKODXPLIEKLAETIVNSFYKETAEHLQSGGEKKK 360
 DB 301 DSTAVALNREDEPKATEIIEPSKODXPLIEKLAETIVNSFYKETAEHLQSGGEKKK 360
 QY 361 ITVFEKISTYTSFCHOLRWYSKRSFKNLGNPQASIAQIIVTVVGLGVIGAIYFGLKND 420
 DB 361 ITVFEKISTYTSFCHOLRWYSKRSFKNLGNPQASIAQIIVTVVGLGVIGAIYFGLKND 420
 QY 421 TGIQNRAGVLFLLTNQCSVSAVELFVEKKLFTHETISGYRVSSYFLGKLLSDLP 480
 DB 421 TGIQNRAGVLFLLTNQCSVSAVELFVEKKLFTHETISGYRVSSYFLGKLLSDLP 480
 QY 481 MTMLPSIIIFTCIYFNLGLKPKADAFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540
 DB 481 MTMLPSIIIFTCIYFNLGLKPKADAFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540
 QY 541 MTICFVFMNIFSGLLVNLTTIASWLSMLQYFSIPRYGFTALQNEFLGONFCGLNATGN 600
 DB 541 MTICFVFMNIFSGLLVNLTTIASWLSMLQYFSIPRYGFTALQNEFLGONFCGLNATGN 600
 QY 601 NPCNVACTGEEYLVKQIDLSPMGLMKNHVALACMIVIFLTAYLKLFLKKYS 655
 DB 601 NPCNVACTGEEYLVKQIDLSPMGLMKNHVALACMIVIFLTAYLKLFLKKYS 655

Search completed: June 6, 2005, 09:40:40
 Job time : 169 secs


```
QY 361 ITVFEKISYTTSPFCHOLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5 420
Db 361 ITVFEKISYTTSPFCHOLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5 420
QY 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLLSDLP 480
Db 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLLSDLP 480
QY 481 MTMLPSIIFTCIYFVNLGLPKADAFVVMFTLMVAVYSSSMALAAAGQSVSVATLL 540
Db 481 MTMLPSIIFTCIYFVNLGLPKADAFVVMFTLMVAVYSSSMALAAAGQSVSVATLL 540
QY 541 MTICFVPMIFSGLLVNLTTIASWLSMLOFSPRIGFTALQHNHFLQGFQPCGLNATGN 600
Db 541 MTICFVPMIFSGLLVNLTTIASWLSMLOFSPRIGFTALQHNHFLQGFQPCGLNATGN 600
QY 601 NPCNYATCTGEEYLVKQIDLSFPMGLMKHVALACMIVIFLTIAYLKLFLKKYS 655
Db 601 NPCNYATCTGEEYLVKQIDLSFPMGLMKHVALACMIVIFLTIAYLKLFLKKYS 655
```

RESULT 2

```
US-09-767-594-1
; Sequence 1, Application US/09767594
; Patent No. 6521635
; GENERAL INFORMATION:
; APPLICANT: Bates, Susan
; APPLICANT: Robey, Robert
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
; FILE REFERENCE: 015280-402100US
; CURRENT APPLICATION NUMBER: US/09/767, 594
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 60/177,410
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human mitochondrion resistance (MXR)/BRCP/ABCP
; OTHER INFORMATION: protein
US-09-767-594-1
```

```
Query Match 99.4%; Score 3331; DB 4; Length 655;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCRKPYE 60
Db 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKGLNALIPTGGGKSSLLDVLAARQPSGLSGDVLINGAPPANFKN 120
Db 61 KEILSNINGIMKGLNALIPTGGGKSSLLDVLAARQPSGLSGDVLINGAPPANFKN 120
QY 121 SGVYVODDVVWGTLTVRENLOFSALRLATMTNHEKNERINRYIOELGLDKVADSKVT 180
Db 121 SGVYVODDVVWGTLTVRENLOFSALRLATMTNHEKNERINRYIOELGLDKVADSKVT 180
QY 181 QFIRGVSGGERRKTSIGMELITDPSILFLDEPTTGDDSTANAVALLKMSKQGRITIF 240
Db 181 QFIRGVSGGERRKTSIGMELITDPSILFLDEPTTGDDSTANAVALLKMSKQGRITIF 240
QY 241 SHQPRYSIFKLFDLSLTLASGRIMFHPQAOEALGYFESAGYHCEAVNNPADPFLDING 300
Db 241 SHQPRYSIFKLFDLSLTLASGRIMFHPQAOEALGYFESAGYHCEAVNNPADPFLDING 300
```

```
QY 301 DSTAVALLNREEDFKATEIIEPSKODKPLIEKLAIEIVNNSFYKETAEHLQSLGGEKKK 360
Db 301 DSTAVALLNREEDFKATEIIEPSKODKPLIEKLAIEIVNNSFYKETAEHLQSLGGEKKK 360
QY 361 ITVFEKISYTTSPFCHOLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5 420
Db 361 ITVFEKISYTTSPFCHOLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5 420
QY 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLLSDLP 480
Db 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLLSDLP 480
QY 481 MTMLPSIIFTCIYFVNLGLPKADAFVVMFTLMVAVYSSSMALAAAGQSVSVATLL 540
Db 481 MTMLPSIIFTCIYFVNLGLPKADAFVVMFTLMVAVYSSSMALAAAGQSVSVATLL 540
QY 541 MTICFVPMIFSGLLVNLTTIASWLSMLOFSPRIGFTALQHNHFLQGFQPCGLNATGN 600
Db 541 MTICFVPMIFSGLLVNLTTIASWLSMLOFSPRIGFTALQHNHFLQGFQPCGLNATGN 600
QY 601 NPCNYATCTGEEYLVKQIDLSFPMGLMKHVALACMIVIFLTIAYLKLFLKKYS 655
Db 601 NPCNYATCTGEEYLVKQIDLSFPMGLMKHVALACMIVIFLTIAYLKLFLKKYS 655
```

RESULT 3

```
US-09-538-092-72
; Sequence 72, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538, 092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratasegFormatter Version 0.9
; SEQ ID NO 72
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YCR011C
US-09-538-092-72
```

```
Query Match 24.9%; Score 835.5; DB 4; Length 1049;
Best Local Similarity 30.5%; Pred. No. 3e-77; Indels 115; Gaps 18;
Matches 222; Conservative 134; Mismatches 257; Indels 115; Gaps 18;

QY 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCRKPYE 60
Db 355 LGSSKSPIRLP-DEDAVNNFLQNEEDTL-----ALTSFENITYVSPSINS-----DGYE 402
QY 61 KEILSNINGIMKGLNALIPTGGGKSSLLDVLAARQPSGLSGDVLINGAPPANFKN 118
Db 403 EYVLANISGIVKRGQILAIMGSGAGAKTLLDILAMKRTGHVSGSIKVGISMDRKSFS 462
QY 119 CNGYVYVODDVVWGTLTVRENLOFSALRLATMTNHEKNERINRYIOELGLDKVADSKV 178
Db 463 KIIGFVODDFFLPTLVETVINSALRLPKALSEAKKARYKVALEIRIIDIORIT 522
QY 179 GTOFIRGVSGGERRKTSIGMELITDPSILFLDEPTTGDDSTANAVALLKMSKQGRITIF 237
Db 523 GNEFDRGISGGERKRSIACELVTSPLVFLDEPTTGDDSNANNVITBCLVRLSSDYNRT 582
QY 238 IIFSHPRYSIFKLFDLSLTLASGRIMFHPQAOEALGYFESAGYHCEAVNNPADPFLDI 297
```

```
Db      583 LVLSTHQPNSIFVLPDKVLVLLSKGEMVYSGNAKKVSEFLRNBEGYICPDNVIADVLIDI 642
Qy      298 -----INGSTAV 305
Db      643 TFEAPQCKRRIRINISDLAECTDNDINTIHQTTFTSSDGTTOREMAHLAHRDEIRS 702
Qy      306 ALNREEDFKATE-----IIEPSKODKPLIEKLAIEIYVNSGFYKETAELHQ-LSGGEKKK 360
Db      703 LLRDEDEVBGTGRGATETIDNTKLLHK-----YDSVYAELESGEIEVLSGDEEN 758
Qy      361 IT--VFKEISYTTSCHQLRWVSQSGFKLLGNPOASIAQIIVTVVLGLVIGAIYFGKNI 418
Db      759 VLNGDLPTQOQAGFLQQLSILNSRSFKWMYRNPKLLNGVLLTILSLFGLTYLVNEN 818
Qy      419 DSTGQNAGVLPFLTTNOCSSVASAVELFVVEKULFIHEIYISGYRVSYFLGKLSL 478
Db      819 DISGQNRKGLFFFLTYGFTFTGLSSPALERIIPIKERNNYSPPLAAYISKIMSEV 878
Qy      479 LPMTMLPSIIFTCIYFVLGLPKADAFVMMFTLMMVAVSASMAIAAGSVSVAT 538
Db      879 VFLRVVPLLSLIYPMGLMKDPAFKICIGILIFMLGISLELITIGIIFEDLANSI 938
Qy      539 LMTICFVPMIIFSGLLV---NLTTIASWLSMLOYFSIPRYGFTALQNEF----- 586
Db      939 ILSVLVLGSLFLSGFLPINTKNITNVA--FKYLNKNSVFYVAVESLLINEVKTMLKEX 996
Qy      587 LGQNP-CRGLNATGNPNPCVATCTGEBEYLVKQI--DLSPLMKKHVALACMIYFLTI 643
Db      997 YGLNTEVPQ-----ATILSTFGFVQNLVFDIK-----ILALENVFLIM 1036
Qy      644 AYKLLFL 651
Db      1037 GYLALKMI 1044
```

RESULT 4
US-09-538-092-1125

```
Sequence 1125: Application US/09538092
Patent No. 6753314
GENERAL INFORMATION:
APPLICANT: Glot, Loic
APPLICANT: Manafeld, Traci A.
TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
FILE REFERENCE: 15966-542
CURRENT FILING DATE: US/09/538, 092
CURRENT FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/127,352
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/178,965
PRIOR FILING DATE: 2000-02-01
NUMBER OF SEQ ID NOS: 1387
SOFTWARE: CuraPatSeqFormatter Version 0.9
SEQ ID NO 1125
LENGTH: 674
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (0)...(0)
OTHER INFORMATION: Polypeptide Accession Number P45844
US-09-538-092-1125
```

Query Match 21.1%; Score 706.5; DB 4; Length 674;

Best Local Similarity 28.4%; Pred. No. 3, 9e-64;

Matches 194; Conservative 155; Mismatches 251; Indels 83; Gaps 23;

```
Qy      3 SGNVVFIFVSGQNTNGPATAISNDL---KAPT-----SEAV-LSRHNCYRKLSGRFP 54
Db      34 SSNMBA---TETDLNGLHKVYDNNITJEAQRFSLSPPRAAVNIERFDLSVYSEGPBW-- 88
Qy      55 CRKPKKEILSINGIMKPG-LNAILGPTGSGKSLDLVLAARKDPSGLSGDVLINGAPR 113
```

```
Db      89 -RKKGYTKLLKGISGRFNSGELVIAIMGSGACKSTLMNIIAGYRE-TGKKGAVLINGLR 146
Qy      114 PAN-FKNSGYVODDVMTGTLTVRENLQFSALRLATMTHEKNERINRYOELGDK 172
Db      147 DLRCKRKSICYMDMLPHLTVQBAWVSHLQOE--KQGRREWKELITLALGLS 204
Qy      173 VADSKVGTQPIRGVSGRKRKTSIGMEDITDPSILFDEPTTGDSSTANAVALLLKMS 232
Db      205 CANTRTGS-----LSGQGRKRLAIALVELYNNPVPWFPEPTSGLDASCFQVSLMKGLA 259
Qy      233 KQGRITIIHQPRYSIFKLFDLSLTLASGRLMFGPAQOALGYFESAGYHCEATYNNPD 292
Db      260 QGSRSIICITIHQPSAKLPFLPDQLVLSQGCQCYRGKVCNLPYRDLGALNCPYHNPAD 319
Qy      293 PFLDIINDSTVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSGFYKETAEL--KAE 349
Db      320 FMEVAVSG-----EYGDNSRLVRAVRGECDSDHKRLGDAEYN 360
Qy      350 ---HQLSGGEKK-KKITVFKEISYTTSCF-----QLRWKRSFKNLLGNPOASI 396
Db      361 PELMRPBEVYVQTRKLGKRDSSMGCHSFSASCLQFCILFKRTFLSIMRDSVLTH 420
Qy      397 AOIIVTVGLVIGAIYGLKNDSTGIONRAGVLPFLTTNOCSSVASAVEL-FVVEKULF 455
Db      421 LRITSHIGIGLIGLILYIGIGNEAKKVLNSGFLPFMSLFLMFAIMPVLTFPLEMGVF 480
Qy      456 IHEIYISGYRVSYFLGKLSLDPMTMLPSIIFTCIYFVLGLPKADAFVMMFTL-M 514
Db      481 LREHLNYSLSKAVYLAKTMD-VFPQIMFPAYCISYVMKTS-QPSDAVRVLPALADT 538
Qy      515 MVAVSASSMALAIAAGQSVSVATLMTICFVPMIIFSGLLVNLTTIASWLSMLOYFSIP 574
Db      539 MSLVAQSLGILGASASTSLQVATFVGPVTALPVLLFSGFPSPTIPLYLQMSYISYV 598
Qy      575 RYGF-----ALQNEFLGQNPCCGLNATGNPNPCVATCTGEBEYLVKQIDLSPMGLM 627
Db      599 RYGFEGVILSIYGLRED-----LHCDIDETCHFOK---SEAILRE-LDVENAKLY 645
Qy      628 KKHVALACMIYFLTIAYLKLF 650
Db      646 LDFIVLGIFFISLRILNAYFLNY 668
```

RESULT 5

```
US-09-989-981A-2
Sequence 2: Application US/09989981A
Patent No. 6821750
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Board of Regents, The University of Texas System
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: Mus musculus
OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-2
```

Query Match 20.7%; Score 693.5; DB 4; Length 652;

```
Qy      20.7%; Score 693.5; DB 4; Length 652;
```

Best Local Similarity 29.0%; Pred. No. 8,4e-63;
Matches 181; Conservative 142; Mismatches 246; Indels 55; Gaps 16;

```
12 VSQNTNGFPATASNDKATFEGAVLSFHNICRYVKLSGSGFLPCRKEVEKEILSINGIMK 67
25 LGGSVYTGTAARS-----LGVHVSYSVSNRVPMWNITSQOQMTQHLVDVLYVE 73
68 NGIMKPG-LNAIIGPTGGKSSLLDVLAAKDPG-LSGDVLINGAP-RPAPNFKNSGVY 124
74 SLYESQIWCILIGSSGSKTLLDALSRLRGTLEGVFNVCGLRDRDQODCSYV 133
125 VQDDVVMGTLTYRENLQFSALRLATMTNHEKNERINRYOELGLDKVADSKVGTQFIR 184
134 LQSDVLSLTVRETLRYTLMALCRSSADP-YNKKEAVMTSELISHVADQMIQSNFG 192
185 GVSGERKRTSIGMELITDPSILFLDEPTGLDSTANAVLLIKRMSKQRTIIFSIHQ 244
193 GISSGERRRVSIAQLQDPKVMMLDEPTTGLDCTANQIVLLVLAELARRDRIVYTIHQ 252
245 PRYSIFKLPDSLTLASGRMLFHGPAQALGYFESAGYHCEAYNPNADFLDIINGDSTA 304
253 PRSELFOHFDKIALITFVGLVFCGTPEMLGFNNCGYPCPEHSNPFDPYMDLTSVDTQ- 311
305 VALNRE-EDFKATEIIEPSSQDKPLIEKLAETVNSSFYKETAKELHOLSGEKKKITY 363
312 -SREREIETRYKRVQMLECAFESDIYHKI-----LENIERARYLKT 351
364 FKESISYTT-----SFCHQLRWVSKRSFKNLGNPOASIAQIITVVLGVLGAIYFGKND 419
352 LPMVPFPTKDPKPMFGKGLGVLRRVTRNLRNKAQAVMLRVONLIMLFIETLIRYQNN 411
420 ST--GIQNRAGVLFELTNQCFSS-VAVELFVEKKLFHEYISGYRVSSYFLGKLS 476
412 TLGAVQDVRGGLYQVLGATPYTGMLNAVNLFPMLRAVSDQESDDGLYHKMQMLAVL- 470
477 DLPMTMLBSITTCIYVFMGLKPKKADAFVMMFTLM--WVAYSASSALAIAGQSV 533
471 HVPFVSIAIVIFSSVGYMTGLPYEARGYSAALAPHLIGELITLVLGLIVQNPNI 530
534 VSVATLMTICFVFMFIISGLVNLTTIASMLWQFSIPRYGFTALQHNEFLQNF-C 592
531 VNSIVALLIS--GLIGSFIRNIQEMPIPKILGIFTQKICCELIVNEFYGLNFTC 588
593 PGINATGNPCNATCTGEEYLVK 616
589 GGSNTSMNLNHPMCAITGVOFIEK 612

RESULT 6
US-09-989-981A-6
; Sequence 6, Application US/09989981A
; Patent No. 6821750
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABC5 and ABC8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 651
; TYPE: PR1
; ORGANISM: Homo sapiens
```

; FEATURE:
; OTHER INFORMATION: human ABC5 (hABC5)
US-09-989-981A-6

Query Match 20.4%; Score 682.5; DB 4; Length 651;

Best Local Similarity 29.2%; Pred. No. 1.2e-61;
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;

```
13 SQGNTNGFPATASNDKATFEGAVLSFHNICRYVKLSGSGFLPCRKEVEKEILSINGIMK 72
21 SQSLEGARATAP---EPHSLGILHASYSVSHRVRMWNITSQOQMTQHLVDVLYVE 77
73 PG-LNAIIGPTGGKSSLLDVLAAKDPG-LSGDVLING-APRPANFKNSGVYQDDY 129
78 SQIWCILIGSSGSKTLLDAMSGRGRAETLEGVFNVCGLRDRDQODCSYVLOSDP 137
130 VQDGLTYRENLQFSALRLATMTNHEKNERINRYOELGLDKVADSKVGTQFIRGVS 189
138 LSSLTVRETLRYTLMALCRGNGP-SFOKKEAVMAELSLSHVADRLIGNVSLGISTG 196
190 ERKRTSIGMELITDPSILFLDEPTGLDSTANAVLLIKRMSKQRTIIFSIHQPRYSI 249
197 ERRRVSIAQLODPKVMMLFDEPTTGLDCTANQIVLLVLAELARRDRIVYTIHQRSEL 256
250 FKLPDSLTLASGRMLFHGPAQALGYFESAGYHCEAYNPNADFLDIINGDSTAVANLR 309
257 FQLEFKATEIIEPSSQDKPLIEKLAETVNSSFYKETAKELHOLSGEKKKITYVFEISY 369
310 EEDFKATEIIEPSSQDKPLIEKLAETVNSSFYKETAKELHOLSGEKKKITYVFEISY 369
313 ERE-----IETSKR---VQMESAYKKSAICHKT-----LKNIERMHLTTLPMVP 356
370 TT-----SFCHQLRWVSKRSFKNLGNPOASIAQIITVVLGVLGAIYFGKNDST- 421
357 KTQDSRGVSKGLGVLRRVTRNLRNKAQAVMLRVONLIMLFIETLIRYQNN 414
422 GIQNRAGVLFELTNQCFSSVAVELFVEKKLFHEYISGYRVSSYFLGKLSDDLIP 480
415 AIQDRVGLLYQFVATPYTGMLNAVNLFPMLRAVSDQESDDGLYHKMQMLAVL-HVP 473
481 MMLPMLBSITTCIYVFMGLKPKKADAFVMMFTLM--WVAYSASSALAIAGQSV 537
474 FSVATMTIFSSVGYMTGLPYEARGYSAALAPHLIGELITLVLGLIVQNPNI 533
538 TLMITICFVFMFIISGLVNLTTIASMLWQFSIPRYGFTALQHNEFLQNF-C 594
534 VALLSIAGV-LVSGFGLNIGQEMPIPKIISYFTQKICCELIVNEFYGLNFTCGSSN 591
595 LNAATGNPCNATCTG 610
592 VSVTNPMACTGQIPIEKTCPG 615

RESULT 7
US-09-989-981A-4
; Sequence 4, Application US/09989981A
; Patent No. 6821750
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABC5 and ABC8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
```


SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 672
TYPE: PR
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-4

Query Match 19.6%; Score 657.5; DB 4; Length 672;
Best Local Similarity 27.2%; Pred. No. 5e-59;
Matches 185; Conservative 141; Mismatches 270; Indels 85; Gaps 21;

20 PPTASNDLKAFET---EGAVLSFNHCYRVKLSGF-----LPCKRVEKEI--- 63
28 PSEEDNSL-YFTYSGQNTLEVRDLTYQVDIASQVWFEQLAQFIRKSHSSQDSCL 86
64 -LSNINIGMKPG-LNAIIGPTGGKSSLDVLAARKDPSSL-SGDVLINGAPRPANF--K 118
87 GIRNLSFKVRSQGMALAIIGSGGRASLDVITGRHGKMKSGQIWMINGQSTPOLVK 146
119 CMSGVYVDDVVMGTLTRENIQFSAALRLATMTNHEKNERINRYIOELGDKVADSKY 178
147 C-VAHVROHDLPLNLTVAETLAFIAQMLPRTFSQAQRKVEDVIAELRLQCANTRY 205
179 GTPFRGVSGGRKRTSIGMELITDPSILFDEPTTGLDSTANAVILLKMKSGQRTI 238
206 GNTYVRGVSQGRRRVSIQVQLMNPGLILDEPTSGLDSTAHNLVTLTLAKGNRLV 265
229 IFSIHQPRYSIFKLPDSLTLLASGRLMFGRPAQOALGPESAGYHCEAYNPADFLDI 298
266 LISLHQPRSDIFRLFDVLMTSGPTIYGAAGQWQYFTSIGHPCPRYSNPADFYVDLT 325
299 NQDSTAVANLRREDFKATEIIEPSKODKFLIEKLAETVYN-----SSPKETKALHQLS 353
326 SDRRS-----KEREVATVEK-----AQSLALFLEKYQGDFDLP--WKAERKELN 369
354 GGEKKKKITVPKEISYTTSS-----FCHOLRWVSKRSPKULLGNPOASIAQIIVTVVLGV 408
370 TSTHTVSLTLTODTDCGTAVELPMIEBPSTLIRQISNDPRDLPTLLHSGEACLMSTL 429
409 ICAIYFGKKNDSGTGIONRAGVLFPL-----TTNOCFSSVSAVELFVVEKKLFTH 457
430 IGFLLYHGAGAKLSMDTALLFMIGALIPNVILDVYSKCHS-----ERSMLYX 479
458 EYISGYRVSSYFLKLSLDLIPMTLPSIIFTCIYVEMLGKPKADAFVVMFTLMNYA 517
480 ELEDGLYTAGPYFFPAKILGE-LPEHCAYVILIYAMPYIYMLNLRPVBELFLHFLVMLIV 538
518 YSASSMALAIAAGOSVSVATLMTICFPMIIFSGLVNLTIASWL--SMLOYFSTPR 575
539 FCCRTMALAASAMLPTEPHMSFFCNALYNSFYLTGFMINLNL--WIVPAVISKLSPLR 596
576 YGFTALQHNPELFGONFCGLNATGNPNCAVYA--TCTGEETVLVQIGIDLSPMGLMKHVALA 634
597 WQFSLMQIQFNGHLTYTQIG-----NFTTSIIGDPMI--SANDLNSHPYALYLYVI 647
635 CMIVIFLITAYIKLFLPKKYS 655
648 GISYGFLLFYLLSLKLIKOKS 668

RESULT 8
US-09-989-981A-8
Sequence 8, Application US/0998981A
Patent No. 6821750
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
APPLICANT: Board of Regents, The University of Texas System

TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-26
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 673
TYPE: PR
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981A-8

Query Match 19.1%; Score 640.5; DB 4; Length 673;
Best Local Similarity 27.2%; Pred. No. 3e-57;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

16 NTNG-----FPATASNDLKAFTEGA--VLSFNHCYRVKLSGF-----LP---- 54
19 DTSGLQDRLFSSESDNSLYFTYSGQNTLEVRDLTYQVDIASQVWFEQLAQFKMPWTSR 78
55 -CRKVEKEILSNINIGMKPG-LNAIIGPTGGKSSLDVLAARKDPSSL-SGDVLINGA 111
79 SCNSCEIGT-ONLSFKVRSQGMALAIIGSGGRASLDVITGRHGKMKSGQIWMINGQ 137
112 PRPANF--KMSGVYVDDVVMGTLTRENIQFSAALRLATMTNHEKNERINRYIOELG 169
138 PSSPOLVKKC-VAHVROHQLPLNLTVAETLAFIAQMLPRTFSQAQRKVEDVIAELR 196
170 LDKVADSKVQTPFRGVSGGRKRTSIGMELITDPSILFDEPTTGLDSTANAVILLIK 229
197 LRQCADTRGMVYVGLSGERRRVSIGVQLMNPGLILDEPTSGLDSTAHNLVTKLS 256
220 RMSKQRTIIEPSIHQPRYSIFKLPDSLTLLASGRLMFGRPAQOALGPESAGYHCEAYNP 289
257 RLAKGNRLVLSLHQPRSDIFRLFDVLMTSGPTIYGAAGQWQYFTSIGHPCPRYSN 316
290 PADFLDIINGDSTAVANLRREDFKATEIIEPSKODKFLIEKLAETVYNSSPK-ETKAE 348
317 PADFYVDLTSIDR-----SREDELATRE--KAQSLALFLEKVRPL--DPLMKAEK-- 366
349 LHQSGGKKKKITVPKEISYTTSPC-----HQLRWVSKRSPKULLGNPOASIA 397
367 -----DLDEDTCVSSVTPLDTNCLPSPTMPCAVOQFTLLIRQISNDPRDLPTLLI 419
398 QIIVTVVLGVYIGALYFGLKNDSTGIONRAGVLFPL-----TTNOCFSSVSAVE 446
420 HGAECALMSMTIGFLYFHHGSIQLSFMDTALLFMIGALIPNVILDVYSKYS----- 473
447 LFVVEKULFIHEBYISGYRVSSYFLKLSLDLIPMTLPSIIFTCIYVEMLGKPKADAF 506
474 -----ERAMLYLEBEGLYTTGPIYPAKILGE-LPEHCAYIILIYGMPTWLANLRGLQEP 528
507 FVMFTLMVAVYSSSMALAIAGOSVSVATLMTICFPMIIFSGLVNLTIASWLS 566
529 LLHFLVWLIVFCRIMALAALALPTPHMASFFCNALYNSFYLTGFMINLSIMTAPA 588
567 WLOYFSIRYGTALQHNPELFGON--CGLNATGNPNCAVATCTGEETVLVQIGIDLSPMG 625
589 WISKVSFLRWCPEGMKIQFSRRRYKMLGULT-----IAVSGDKIL--SAMELDSVP 639
626 LMKHVALACMIVIFLITAYIKLFLPKK 653
640 LYALYIYIGSGGMVLYVSLRPIKQ 667

RESULT 9
US-09-614-912-138

```
/ Sequence 138, Application US/09614912
/ Patent No. 6677502
/ GENERAL INFORMATION:
/ APPLICANT: Allen, Steve
/ APPLICANT: Rafalski, Antoni
/ APPLICANT: Orozco, Buddy
/ APPLICANT: Miao, Gou-Hau
/ APPLICANT: Famodu, Omolayo O.
/ APPLICANT: Lee, Jian Ming
/ APPLICANT: Sakai, Hajime
/ APPLICANT: Weng, Zude
/ APPLICANT: Caimi, Perry G
/ APPLICANT: Anderson, Shawn
/ TITLE OF INVENTION: Plant Metabolism Genes
/ FILE REFERENCE: B01378 US NA
/ CURRENT APPLICATION NUMBER: US/09/614,912
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: 60/143,401
/ PRIOR FILING DATE: 1999-07-12
/ PRIOR APPLICATION NUMBER: 60/143,412
/ PRIOR FILING DATE: 1999-07-12
/ PRIOR APPLICATION NUMBER: 60/146,650
/ PRIOR FILING DATE: 1999-07-30
/ PRIOR APPLICATION NUMBER: 60/170,906
/ PRIOR FILING DATE: 1999-12-15
/ PRIOR APPLICATION NUMBER: 60/172,959
/ PRIOR FILING DATE: 1999-12-21
/ PRIOR APPLICATION NUMBER: 60/172,946
/ PRIOR FILING DATE: 1999-12-21
/ NUMBER OF SEQ ID NOS: 204
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 138
/ LENGTH: 617
/ TYPE: PRT
/ ORGANISM: Zea mays
US-09-614-912-138

Query Match          16.9%; Score 567; DB 4; Length 617;
Best Local Similarity 28.0%; Pred. No. 1,2e-49;
Matches 169; Conservative 117; Mismatches 224; Indels 94; Gaps 22;
```

```
37 LSFHNICRYVKLKGFLPCRKPEK-----EILSNINGIMKPG-LNALIGPTGG 85
11 MSFDNVVYVDM-----PAEMKQGVQDDRLQLLRVTGSGFRGVTALMGVAG 61
86 KSLLDVLAARKDPSGLSGDVLINGAPR-PANFKNSGYVQDDVWMTLTIVRENLO 144
62 KTLMDVLAARKDPSGLSGDVLINGAPR-PANFKNSGYVQDDVWMTLTIVRENLO 121
145 ALRLATMTNHE-----KNERINRVIQELGLDVAADSKVGTQFIRGVSGERKTSIG 199
122 FLRLPGKIGDQELTDDIKMOPVDEVMELVELDNKRDALVGLPGITGSTQRKRLT 161
200 LITDPSILFLDEPTGDSSTANAVILLKRMKSGKRTIIFSHOPRYSIFKLPDSL 259
182 LVNAPSPITFMEDEPTSGIDARAALIVMRTVNTVDTGRVCTIHOPSIDIFES 241
260 -AGRLMFGP-----AGEALGYFESA-----GYHCEAYNNPADFLDIINDST 306
242 KRGQVYISGKLGNSQKMEYFEALIPGVKIKDKY-----NPATMLEV---SSV 292
307 LNREEDPKATEITEIPSKQDPLIEKLAIEIVYNSGFYKETAELHQLSGEKKK 366
293 VRLKMD-----AKYETSDLYKONKLVNQLSQPBGTSDLVF-P 332
367 ISYTSFCHQLRWVSKSPKLLGNPOASIAQIIVTVVLGLVIGALYFGE---K 423
333 TEVSOSTIGQFKACIMKQMLTYWSPDYNVRYVSFTLLVALLGSI FMRIGT 392
424 QNAGVLF-----FLTNOCSSVSAYVELFVVEKFLFHEYISGYRYSIFLGL 479
393 GNVIGAMVYAVMFIQINNCG-STVQPV--VSIERTVFERRAAGNYSAMPALIA 448
```

```
Qy 480 PNTMLPSIIFTCIVYEMLGKPKKADA-----FFVMMFTLMMVAYSSAMALAIAG 535
449 PVYVQOTYYTLLIVAMSGFQMTAVKFEWFFISYSPFLYFTY-YGMAVVISPNH---E 504
Qy 536 VATLMTICFVEMITSSGLLVNLTITIASWLSMLOYSI-----PRGCFALQNEFL 590
505 VASIFAAFPSSLENLFSGFFIPRPIRPGW--WIMYVICPLAWTVYGLIVTOYDLE 562
Qy 591 FCPG 594
563 SVPG 566
Db
```

RESULT 10

```
US-09-614-912-140
/ Sequence 140, Application US/09614912
/ Patent No. 6677502
/ GENERAL INFORMATION:
/ APPLICANT: Allen, Steve
/ APPLICANT: Rafalski, Antoni
/ APPLICANT: Orozco, Buddy
/ APPLICANT: Miao, Gou-Hau
/ APPLICANT: Famodu, Omolayo O.
/ APPLICANT: Lee, Jian Ming
/ APPLICANT: Sakai, Hajime
/ APPLICANT: Weng, Zude
/ APPLICANT: Caimi, Perry G
/ APPLICANT: Anderson, Shawn
/ TITLE OF INVENTION: Plant Metabolism Genes
/ FILE REFERENCE: B01378 US NA
/ CURRENT APPLICATION NUMBER: US/09/614,912
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: 60/143,401
/ PRIOR FILING DATE: 1999-07-12
/ PRIOR APPLICATION NUMBER: 60/143,412
/ PRIOR FILING DATE: 1999-07-12
/ PRIOR APPLICATION NUMBER: 60/146,650
/ PRIOR FILING DATE: 1999-07-30
/ PRIOR APPLICATION NUMBER: 60/170,906
/ PRIOR FILING DATE: 1999-12-15
/ PRIOR APPLICATION NUMBER: 60/172,959
/ PRIOR FILING DATE: 1999-12-21
/ PRIOR APPLICATION NUMBER: 60/172,946
/ PRIOR FILING DATE: 1999-12-21
/ NUMBER OF SEQ ID NOS: 204
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 140
/ LENGTH: 1296
/ TYPE: PRT
/ ORGANISM: Oryza sativa
US-09-614-912-140
```

Query Match 16.0%; Score 537; DB 4; Length 1296;
Best Local Similarity 27.5%; Pred. No. 5.5e-46;
Matches 162; Conservative 115; Mismatches 230; Indels 82; Gaps 22;

```
Qy 37 LSFHNICRYVKL-----KSGFLPCRKPEKELISNNGIMKPG-LNALIGPTGG 90
690 MSFDNVVYVDMPEAKKQGVQDDRLQLLRVTGSGFRGVTALMGVAGKTTLM 745
91 DVLAAKRDPSGLSGDVLINGAPR-PANFKNSGYVQDDVWMTLTIVRENLO 149
746 DVLAKRGTGVLGDMRISGPNQGTFAISGCEQNDIHSPOVTVRESLIYSAFLR 805
150 TMTNHE-----KNERINRVIQELGLDVAADSKVGTQFIRGVSGERKTSIG 204
806 EKIGDQELTDDIKIQVDEVMELVELDNKDALVGLPGITGSTQBQRKRLTIV 865
205 SILFDEPTGDSSTANAVILLKRMKSGKRTIIFSHOPRYSIFKLPDSL 263
866 SILFMEDEPTSGIDARAALIVMRTVNTVDTGRVCTIHOPSIDIFEAFDEL 925
264 LMFHGP-----AGEALGYFESA-----GYHCEAYNNPADFLDIINDST 311
```

```
Db 926 VYISQGLGRNSQKMEYFEALPGVPRKIKDKY-----NPATMWLEV---SSVAAEVRLNM 976
Qy 312 DPKAETIEIPSKODKPLIEKLAETIYVNSFFYKETAELHQLSGGKKKIITYFKELSYTT 371
Db 977 DF-----ABEYKTSIDLKQNKVNLVQLSPBEGTSDLHF-PTKXGQ 1016
Qy 372 SFCCHLRWVSKSPKMLGNPQASIAQIIVTVVLGLVIGAIYF--GLK-NDSTGIONRAG 428
Db 1017 STIGCFRACLKQWMLTYRSPDYNLVRFSEFTLFTALLLGTITFKKGTGKGNANSLRWIG 1076
Qy 429 VLF-----FLTINQCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGLSLDLPMTWL 484
Db 1077 AMYTAVMFIGINNC---ATVQPIVSIERTVFRERAAQWVSAMPYALQVME-IPYVFW 1132
Qy 485 PSIIITCIYFVFLGKPKKADA---FVVMFTLMVAVSASSMALAIAGQSVSVATLL 540
Db 1133 QTAAYVTLVYAMMSQMTAKFFWFFVVSFSLYFTY-YGMWTVASISPH--EVAALF 1188
Qy 541 MTICVFMMIFSGLLVNLTTIASWLSWLOYSIPR---YGFALQHNE 585
Db 1189 AAAYFSLENLFSRFFIIPRPIKRWIM-YWLCPLAMTVYGLITVQYGD 1236

RESULT 11
US-09-248-796A-20212
; Sequence 20212, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20212
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (740)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-20212

Query Match 15.4%; Score 516; DB 4; Length 740;
Best Local Similarity 26.7%; Pred. No. 3.3e-44;
Matches 169; Conservative 106; Mismatches 245; Indels 112; Gaps 21;

Qy 16 NNGCPATASNDLKAFTESAVLSFNICVRVLSKGFPCRKRVK---ELISNINGIMK 72
Db 135 NDSDIQPTVNALMFTTEAT-----NKLK-----KPDSSKTFDILKSDALNR 178
Qy 73 PG-LNALIGPTGGKSSLDVLAAR-----KDP-----SGLS-----GDVLINGA 111
Db 179 PGEIIVVIGRPAGSGSTLKTIAVMTYGFHIGKEQITDGLSPHDIERNHGVDI----- 234
Qy 112 PRPAFKNGSVVODVVMGTLTVRENLOFSAALRLATTMTNHEKNERIN----- 163
Db 235 -----YSAETDVHFPRLSVGDTLEFAARL-----TPQNRGEGIDRETYAKMA 278
Qy 164 ---VIGELGLDKVADKVGQFIRGVSGGRKRTSIMGELITDPSILFDEPTGDSSTA 221
Db 279 SVYMAITYGSHTRNTNNGNDPVRGVSGBGRKVSIAEASLQICWMDNATRCUDSAT 338
Qy 222 NAVLLLKRMKSK-OGRTIIFSIHOPRYSIFKFLPDSLTLTLLASGRLMFHGPACALGYFESA 280
Db 339 LEFIALKTSATILDTTLPALAIYQCSQDAVELFDVNVVLYEGYQIFPGKASAKRYFENN 398
```

```
Qy 281 GYHCEAVNNPADPFLDITNGDSTAVALNREDDFKATEIIEPSKODK--PLIEKLAETIYVN 338
Db 399 GWKCKQROQTADPFLSLTN-----PAEREP-----LPGVEDKVPRTAGEFEFWMGN 444
Qy 339 SSGFYKETAELHQL-----SGGKKKIITYFKELSYTT-----YTSFCCHLRWVSKS 384
Db 445 PEVEALTEIDEYCEVCEGRSNTGETYRSHVAKOSSNTRPSPPTVSPFQKRVYIARN 504
Qy 385 FKNLGNPQASIAQIIVTVVLGLVIGAIYFGKANDSTGQNAQVLFPLTNQCFSS-VS 443
Db 505 FLRMKGDHSIHIFSFQGLWMLIISVFNYSQTFGSFYGAAMIFAVLFNAPSSSLE 564
Qy 444 AVELF-----VVEKKLFIEHYISGYRVSSYFLGLSLDLPMTWLPSIIFTCIYVFMGL 499
Db 565 IMSLEAPRIYEK---HKYA-FIRPADALASIISE-LPVKLMNSFNVPFPMVNF 618
Qy 500 KPKADAFVVMFTLMVAVSASSMALAIAGQSVSVATLLMTICFVPMIFSGLLVNL 559
Db 619 RNPGRFFPYWLMCTWCFWVSHLFRSIGAVTSISGAMTPATVLLANVITGFIPIPR 678
Qy 560 TIASWLSWLOYSIPRYGFTALQHNEFLGQNF 591
Db 679 SWLGSRWIMYINIPGVYFESLMVNEFHGREF 710

RESULT 12
US-09-248-796A-20772
; Sequence 20772, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20772
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20772

Query Match 15.3%; Score 514.5; DB 4; Length 518;
Best Local Similarity 29.4%; Pred. No. 2.6e-44;
Matches 161; Conservative 112; Mismatches 217; Indels 57; Gaps 22;

Qy 121 SGYVVO-DVVMGTLTVRENLOFSAALRLATTMTNHEKNER-INRVIQELGLDKVADSKV 178
Db 15 SAYGIQHSPLIEQLTVARETLTYQAKL--PLDQHKFPIITINKLITGVLVDCADTLI 72
Qy 179 GTQFIRGVSGGRKRTSIMGELITDPSILFDEPTGDSSTANAALLKRMKSKGR-T 237
Db 73 GSEYVKGISGGRKRVSAIQLSKPKVLPDEPTSGDSSYAEVITLLGLLAKENMTT 132
Qy 238 IIFSIHOPRYSIFKFLPDSLTLTLLA-SGRLMFHGPACALGYFESAGYHCEAVNNPADPFLD 296
Db 133 IILTHQPSQULFYFGSLLLGRCQKVIYDTSIGIYVLESLOGYNNPEGHNIDYILD 192
Qy 297 IINGDSTAVALNREDDFKATEIIEPSKODKPLIEKLAETIYVNSFFYKETAELHQLSGGE 356
Db 193 LISRMNDKQGLER--RVAEI--SYQANSIKKLGCS---TATPLQEI-IDLPQY---- 240
Qy 357 KKKKITVKEISYTTSPCHLRWVSKSPKMLGNPQASIAQIIVTVVLGLVIGAIYFGL 416
Db 241 YYORLPIF--ITFPPIFRRL--LTSYRAKQVNVINRACQ-----TIFLAIYHTLYFTPL 290
Qy 417 KNDSTGIONRAGVLFPLTNQCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGLKLS 476
```

Db 291 RNTQEGISNRLGVOEVLNLYFAGLINNITLYPERNLFOEYRDGIVGTGERLSYLIN 350
QY 477 DILPMMLPSIIITFCIVYFMGLKPKRDAFFVMMFTLMVAIYSSAALAIAGQSVSV 536
Db 351 E-LPTETVPCGFPAALVFACGLPRTPOMFAMFGTGVINCESIGITVNSLFTMGV 409
QY 537 ATLMTICFVPMFMMIFSGILVNLFTIASWLSMLQYFSIPRYGFTALQHNELGONFCPLN 596
Db 410 ATNVLS-TFVSLAIFMCGTMSL-HMPCFGKINPISMKYAVAI CANLGFRKNSF----- 462
QY 597 ATGNPNPN--YATC---TGEE---YLKQIGIDLSPMGLMKNHVALACMIVIFLTAYLK 647
Db 463 -----KCNSEAADCLLTGEBDVLSYNMKNLGPVGL-----IGC-LVIYRVAILLS 510
QY 648 LFLFKKY 654
Db 511 IYRVK 517

RESULT 13

US-09-538-092-413
; Sequence 413, Application US/09538092
; Patent No. 675314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538, 092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; SOFTWARE: CurataseqFormatter Version 0.9
; SEQ ID NO 413
; LENGTH: 1411
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Y11013C
US-09-538-092-413

Query Match 14.6%; Score 488.5; DB 4; Length 1411;
Best Local Similarity 25.1%; Pred. No. 7.2e-41;
Matches 170; Conservative 137; Mismatches 272; Indels 99; Gaps 24;

QY 32 TEGAVISFNHCYRVKLSGFLPCRKVEKEILSNINGIMKPGIINALIGPTGGKSLD 91
Db 746 TQGHVISMKNINVTIG-----DKKLINDAGYISSGTLALMGSGAGKTTLLN 793
QY 92 VLAARDPSGLSDVILNGAP--RPAFKNSGVVDDVVMGTLVYREMLQFSAALRLA 149
Db 794 VLSQRFESGVVTELLIDGQPLTNIDAFRSIGFVQGDVHLELLTYRESLEISCVLR-- 851
QY 150 TWTNHEKERNIRVIOELGLDKVADSKVGTQPIRGVSGGERKRTSIGMELTDPG--ILF 208
Db 852 -----GQDPRDYLGVNSLRLP-----SEKLVDLSPTQKLLSIGVELVTKPSILLF 900
QY 209 LDEPTTGLDSTANAVILLKMSKOGRTIIFSIHQRYSIIFKLFDLSLTL--ASGRLMFH 267
Db 901 LDEPTTGLDSEALVTVQFLKLSMOGALICTIHQSKSVISYFDNIYLLKRGGEVVF 960
QY 268 GPQOELAGYESAGYHC---EAYNPPADFLDIINGDSTAVALNREDPFALEIIP-- 321
Db 961 GSPJNACDLYVAIDRLITPREMDNPADYIVDVGSSTNIIPMDAEKPTSSKIDPEVSY 1020
QY 322 SKQDKPLIEKLAIEIVYNSFFYKETKALHQLSGGKKKKITVFKEISYTTSPFCHQLRWVS 381

Db 1021 HKQSDSI--NMAELWQSPBEKRVVADLLLEBEARKSGVDTTYSVWSPSYEQIKLIT 1078
QY 382 KRSFKNLGNPQASIOIITVTVGLVIGAIYGLKNDNSGIONRAGVLEFLTNOCFSS 441
Db 1079 KROYICTKMDTVYPAKYALNAGAGLFIGSFRTKHNINGLDALFLCMMI-----CVSS 1135
QY 442 --VSARELVPEVEKKLFIHEIYISGYRVSSYFLCKLSDL----LPMTMLPSIIFTCIVF 495
Db 1136 PLINQODKALQSK---EVIYAREARSNTFWTVLLIAQIVELPLAISSTFLFLCCYF 1192
QY 496 MGLKPKRA--DAFVMMFTLMVAIYSSAALAIAGQSVSVATLMTICFVPMFMMIFSG 553
Db 1193 CCGFETSARVAGVFNLYNLIFSMTYLSFGIMLLYSAPD--LQTAAVFAFLYSFTASFCG 1250
QY 554 LVLNLTIASWLSMLQYFSIPRYG-----TALQHN-----EFLGQ 589
Db 1251 VMQPSLPRRF--WTMWRVSPYTYIEFVLSLLHDREVNCGSTSENVQPMVGQCGQ 1308
QY 590 NFGPGLNATG-----NNP---CNYPATCT--GEELYVKGIDLSPMGLMKNH-----VA 632
Db 1309 FMKPFIDFGCKLHINNTYTVCAVCMYTVGDDEFLAQE--NMSYHHWRNRFGEVWFVCFN 1366
QY 633 LACMIVIFLTIAVLKLLP 650
Db 1367 IAAMFVGFY-LTYIKKIW 1383

RESULT 14

US-09-270-767-43468
; Sequence 43468, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 43468
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-43468

Query Match 14.4%; Score 483.5; DB 4; Length 520;
Best Local Similarity 29.4%; Pred. No. 4.4e-41;
Matches 128; Conservative 78; Mismatches 162; Indels 67; Gaps 8;

QY 60 EKEILSNINGIMKP-GLNALIGPTGGKSLDVLAAKDPGSLGVDILNGAPRPANFX 118
Db 90 ERIILDSVGSFVPCVLAIMGSSGKTTLLDCLSGQRIID--SGSVFLNREPLTYKWR 147
QY 119 CNSGYVDDVVMGTLVRENLQFSAALRLATMTNHEKERNIRVIOELGLDKVADSKV 178
Db 148 RRIQYVIGIEIFPPQLTLETVYVYALLRLBESMPAEKMRQVDHILEALELCCQQTKE 207
QY 179 GTFPIGVSGGERKRSIGMELTDPESILELDEPTTGLDSTANAVILLKMSK--QGR 237
Db 208 GDVNLNGLSGGEKRRNIAEELTNPLMLLDEPTGLDSHSAISLMKVLKRYAQLEQKT 267
QY 238 IIFSIIHQPRYSIFKLFDLSLTLASGRLMFPGQOELAGYESAGYCEAYNPPADFLD-- 296
Db 268 IYVSVIQPSQOMHMDKLLHQLGRTAYGDVQNYIRHREDYGVITIKRYNADPFLBQ 327
QY 297 -----LINDSTAVALNR-----BEDFKATEIIPSKQDKPLIEKLAIEIV 337
Db 328 LKSHPDIREKLFIYAKESHGNYLNRCITSSHHNVQSVGAKGKKQADSLIDDIINNY 387
QY 338 N-----SSFYETKALHQLSGGK 357
Db 388 NQSRNHHQYENLHHTNSGCRVEEDEEAQHLWCAADQSNSFSCASSDCHSYSGSG 447

Qy 358 KKKITVFKKISYTSFCHQLRWVSKRSPKULLGNPQ-ASIAQIYTVVLGLVIGAIYFGL 416
Db 448 PCHSADDDWLSTYPTSFHTQFCVLSNRFRE--AKPRMLSKLWMPFOTIGLALMAGAIWFOQL 505
Qy 417 KNDSTGIONRAGVLP 431
Db 506 PRTEEFHLDLQGMF 520

RESULT 15

US-09-248-796A-20770
; Sequence 20770, Application US/09248796A
; Patent No. 6741137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20770
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20770

Query Match 13.8%; Score 463; DB 4; Length 300;

Best Local Similarity 41.7%; Pred. No. 2.4e-39; Matches 106; Conservative 49; Mismatches 87; Indels 12; Gaps 6;

Qy 61 KEILSNINGIMKPG-LNAILGPTGGKSLDLVLAARKDP--SGLSGDLINGAPRPAN- 116
Db 27 KTLIDDIYGSVAGEMLAIMGSGCKSTLNLVLAVRTSPRSTLEGIFINNERATLNK 86
Qy 117 FKCNAGYVQDDVNGTLTVRENLOPSAALRLATTMTNHEKNERINRYIQELGDKVADS 176
Db 87 IKQLSSVVEQEDSLIGSLVLETVDSAP--AGIDKAHKKELVSKTKISGLEQAML 143
Qy 177 KVGTFIVGSGGERKRTSIGMELITDPBILFLDEPTGLDSTANAVLLIKRMSK--QG 235
Db 144 KIGTIOKISGQKRVSIASQIITSPIFLDEPTSGLDVVASREVISTIKTIKAKREN 203
Qy 236 RTIIFSIHQPRYSIFKLFDLTLASGRIMFHPAQOALGYFESAGYHCEAYNPPADPFL 295
Db 204 MWIISIHQPSYTFELFDKWMFLSKGRVYNGAVSNVYKTFNSIGHTMPYINPAEYVL 263
Qy 296 DIIN---GDSTAV 305
Db 264 DLINTDFQGDSSVL 277

Search completed: June 6, 2005, 09:45:19
Job time : 45 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using bw model

Run on: June 6, 2005, 09:43:50 ; Search time 141 Seconds
(without alignments)
1605.818 Million cell updates/sec

Title: US-09-961-086-1
Perfect score: 3352
Sequence: 1 MSSSVVEFIPVSGQNTNGF.....MIVIFLTAVLKLFKKYS 655

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1465611 seqs, 345679903 residues
Total number of hits satisfying chosen parameters: 1465611

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3352	100.0	655	10 US-09-961-086-1	Sequence 1, Appl
2	3352	100.0	655	15 US-10-405-806-13	Sequence 13, Appl
3	3346	99.8	655	9 US-09-981-353-35	Sequence 35, Appl
4	3346	99.8	655	14 US-10-120-687-61	Sequence 61, Appl
5	3346	99.8	655	15 US-10-405-806-2	Sequence 2, Appl
6	3346	99.8	655	17 US-10-874-706-24	Sequence 24, Appl
7	3338	99.6	655	9 US-09-866-866A-27	Sequence 27, Appl
8	3331	99.4	655	9 US-09-866-866A-10	Sequence 10, Appl
9	3331	99.4	655	14 US-10-090-455-5	Sequence 5, Appl
10	3053.5	91.1	604	9 US-09-745-763-197	Sequence 197, Appl
11	2757	82.2	657	9 US-09-866-866A-14	Sequence 14, Appl
12	835.5	24.9	1049	15 US-10-369-493-1520	Sequence 1520, Ap
13	812	24.2	663	13 US-10-108-605-245	Sequence 245, Ap

14	812	24.2	687	17 US-10-732-923-1785	Sequence 1785, Ap
15	812	24.2	687	17 US-10-732-923-1786	Sequence 1786, Ap
16	812	24.2	687	17 US-10-732-923-1784	Sequence 1784, Ap
17	808	24.1	687	17 US-10-732-923-1775	Sequence 1775, Ap
18	806	24.0	687	17 US-10-732-923-1791	Sequence 1791, Ap
19	805	24.0	1084	15 US-10-424-599-242078	Sequence 242078, A
20	805	24.0	1101	15 US-10-425-114-63125	Sequence 63125, A
21	795.5	23.7	679	15 US-10-369-493-3799	Sequence 3799, Ap
22	793	23.7	679	17 US-10-732-923-1781	Sequence 1781, Ap
23	790	23.6	692	17 US-10-732-923-1777	Sequence 1777, Ap
24	777.5	23.2	679	17 US-10-732-923-1588	Sequence 1588, Ap
25	773.5	23.1	695	15 US-10-424-599-176182	Sequence 176182, A
26	772	23.0	679	17 US-10-732-923-1760	Sequence 1760, Ap
27	772	23.0	679	17 US-10-732-923-1783	Sequence 1783, Ap
28	760	22.7	666	17 US-10-732-923-1592	Sequence 1592, Ap
29	760	22.7	706	17 US-10-732-923-1789	Sequence 1789, Ap
30	754.5	22.5	677	17 US-10-732-923-1794	Sequence 1794, Ap
31	751	22.4	669	17 US-10-732-923-1773	Sequence 1773, Ap
32	751	22.4	695	17 US-10-732-923-1778	Sequence 1778, Ap
33	746.5	22.3	819	15 US-10-425-114-54421	Sequence 54421, A
34	745.5	22.2	673	17 US-10-732-923-1779	Sequence 1779, Ap
35	744	22.2	669	17 US-10-732-923-1774	Sequence 1774, Ap
36	743.5	22.2	1528	17 US-10-732-923-1678	Sequence 1678, Ap
37	743	22.2	1328	17 US-10-732-923-1677	Sequence 1677, Ap
38	739	22.0	612	16 US-10-437-963-185375	Sequence 185375, A
39	736.5	22.0	709	17 US-10-732-923-1682	Sequence 1682, Ap
40	730.5	21.8	1520	17 US-10-732-923-1582	Sequence 1582, Ap
41	728.5	21.7	711	16 US-10-437-963-123860	Sequence 123860, A
42	727	21.7	832	15 US-10-267-502-303	Sequence 303, Ap
43	723.5	21.6	700	16 US-10-437-963-180018	Sequence 180018, A
44	716	21.4	610	17 US-10-732-923-1765	Sequence 1765, Ap
45	715.5	21.3	1393	17 US-10-732-923-1681	Sequence 1681, Ap

ALIGNMENTS

RESULT 1
US-09-961-086-1
Sequence 1, Application US/09961086
General Information:
Publication No. US20030036645A1
Applicant: UNIVERSITY OF MARYLAND, BALTIMORE
Applicant: ROSS, Douglas D.
Applicant: DOYLE, L. Austin
Applicant: ABRUZZO, Lynne
Title of Invention: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
File Reference: EP19376-019
Current Application Number: US/09/961,086
Current Filing Date: 2001-09-21
Prior Application Number: US 60/073,763
Prior Filing Date: 1998-02-05
Prior Application Number: PCT/US99/02577
Prior Filing Date: 1999-02-05
Number of Seq ID NOS: 7
Software: PatentIn Ver. 2.1
Seq ID NO 1
Length: 655
Type: PRT
Organism: Homo sapiens
US-09-961-086-1
Query Match 100.0%; Score 3352; DB 10; Length 655;
Best Local Similarity 100.0%; Pred. No. 2.2e-288;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSSSVVEFIPVSGQNTNGPATASNDIKAFEGAVLSFHNICVYKXKSGSLPCKRPPVE 60
DB 1 MSSSVVEFIPVSGQNTNGPATASNDIKAFEGAVLSFHNICVYKXKSGSLPCKRPPVE 60
QY 61 KEILSININGIMKPGINALILGPTGGGKSLDVLAAKDPGSLGSDVLINGAPRPANPKCN 120

Db 61 KEILSNINGIMKGLNALIIGPTGGKSSLLDVLAAKDPGSLGVDLNGAPPANFKCN 120
Qy 121 SGYVDDVVMGTLTYRENLOFSALRLATMTNHEKNERINRYIQELGLDKVADSKVGT 180
Db 121 SGYVDDVVMGTLTYRENLOFSALRLATMTNHEKNERINRYIQELGLDKVADSKVGT 180
Qy 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
Db 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
Qy 241 SIHQPRYSIFKLPDSLTLLASGRLMFHPGPAQALGFESAGYHCEAYNPADPFLLDING 300
Db 241 SIHQPRYSIFKLPDSLTLLASGRLMFHPGPAQALGFESAGYHCEAYNPADPFLLDING 300
Qy 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
Db 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
Qy 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420
Db 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420
Qy 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Db 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Qy 481 MTMLPSIIIFTCIYFMGLKPKADAFVMMFTLMMVAYSSSMALAIAGQSVSVATLL 540
Db 481 MTMLPSIIIFTCIYFMGLKPKADAFVMMFTLMMVAYSSSMALAIAGQSVSVATLL 540
Qy 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNEFLGQFCGLNATGN 600
Db 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNEFLGQFCGLNATGN 600
Qy 601 NPCNVAATCGEEYLVKQIDLSPWGLMKHVALACMIVIFLTAYIKLFLKXYS 655
Db 601 NPCNVAATCGEEYLVKQIDLSPWGLMKHVALACMIVIFLTAYIKLFLKXYS 655

RESULT 2
US-10-405-806-13
; Sequence 13, Application US/10405806
; Publication No. US2003023262A1
; GENERAL INFORMATION:
; APPLICANT: KOMATANI, HIDEYA
; APPLICANT: HARAI, YOSHIKAZU
; APPLICANT: KOTANI, HIDEHITO
; APPLICANT: NAKAGAWA, RINAKO
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF
; FILE REFERENCE: 234985USOCNT
; CURRENT APPLICATION NUMBER: US/10/405, 806
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: PCT/JP01/08112
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP2000-303441
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ABCG2 482mutant sequence
US-10-405-806-13

Query Match 100.0%; Score 3352; DB 15; Length 655;
Best Local Similarity 100.0%; Pred. No. 2, 2e-288;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSSSNVEVFIPIVSGQNTNGFPATASNDLKAFTBGAVLSFNICYRYVLSKSGFLPCRKPE 60
Db 1 MSSSNVEVFIPIVSGQNTNGFPATASNDLKAFTBGAVLSFNICYRYVLSKSGFLPCRKPE 60

Qy 61 KEILSNINGIMKGLNALIIGPTGGKSSLLDVLAAKDPGSLGVDLNGAPPANFKCN 120
Db 61 KEILSNINGIMKGLNALIIGPTGGKSSLLDVLAAKDPGSLGVDLNGAPPANFKCN 120
Qy 121 SGYVDDVVMGTLTYRENLOFSALRLATMTNHEKNERINRYIQELGLDKVADSKVGT 180
Db 121 SGYVDDVVMGTLTYRENLOFSALRLATMTNHEKNERINRYIQELGLDKVADSKVGT 180
Qy 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
Db 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240
Qy 241 SIHQPRYSIFKLPDSLTLLASGRLMFHPGPAQALGFESAGYHCEAYNPADPFLLDING 300
Db 241 SIHQPRYSIFKLPDSLTLLASGRLMFHPGPAQALGFESAGYHCEAYNPADPFLLDING 300
Qy 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
Db 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
Qy 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420
Db 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420
Qy 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Db 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLSDDL 480
Qy 481 MTMLPSIIIFTCIYFMGLKPKADAFVMMFTLMMVAYSSSMALAIAGQSVSVATLL 540
Db 481 MTMLPSIIIFTCIYFMGLKPKADAFVMMFTLMMVAYSSSMALAIAGQSVSVATLL 540
Qy 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNEFLGQFCGLNATGN 600
Db 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNEFLGQFCGLNATGN 600
Qy 601 NPCNVAATCGEEYLVKQIDLSPWGLMKHVALACMIVIFLTAYIKLFLKXYS 655
Db 601 NPCNVAATCGEEYLVKQIDLSPWGLMKHVALACMIVIFLTAYIKLFLKXYS 655

RESULT 3
US-09-981-353-35
; Sequence 35, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981, 353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 35
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
US-09-981-353-35

Query Match 99.8%; Score 3346; DB 9; Length 655;
Best Local Similarity 99.8%; Pred. No. 7, 5e-288;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSSSNVEVFIPIVSGQNTNGFPATASNDLKAFTBGAVLSFNICYRYVLSKSGFLPCRKPE 60
Db 1 MSSSNVEVFIPIVSGQNTNGFPATASNDLKAFTBGAVLSFNICYRYVLSKSGFLPCRKPE 60
Qy 61 KEILSNINGIMKGLNALIIGPTGGKSSLLDVLAAKDPGSLGVDLNGAPPANFKCN 120


```
Db      61 KEILSININGIMKPGNALIIGPTGGKSSLLDVLAAKDPBSGLSGDVLINGAPRPAKFKN 120
Qy      121 SGVVODDVVNGTLTVRENLOFSALRLATTTMTHNEKNRINRVIOELGLDYADSKVGT 180
Db      121 SGVVODDVVNGTLTVRENLOFSALRLATTTMTHNEKNRINRVIOELGLDYADSKVGT 180
Qy      181 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLLKRMKSOGRTIIF 240
Db      181 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLLKRMKSOGRTIIF 240
Qy      241 SHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADPFLDIING 300
Db      241 SHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADPFLDIING 300
Qy      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSPFYKETAELHOLSGGEKKKK 360
Db      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSPFYKETAELHOLSGGEKKKK 360
Qy      361 ITVPKEISYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIITVVLGLVIGALYFGLKXDS 420
Db      361 ITVPKEISYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIITVVLGLVIGALYFGLKXDS 420
Qy      421 TGIORAGVLPFLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLLSDLLP 480
Db      421 TGIORAGVLPFLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLLSDLLP 480
Qy      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Db      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Qy      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Db      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Qy      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 655
Db      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 655
```

```
RESULT 4
US-10-120-687-61
; Sequence 61, Application US/10120687
; Publication No. US20030082155A1
; GENERAL INFORMATION:
; APPLICANT: Massachusetts General Hospital
; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating
; FILE REFERENCE: 3284/1235B
; CURRENT APPLICATION NUMBER: US/10/120,687
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: US60/169082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 09/963,875
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/215109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/238880
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/731261
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-120-687-61
```

Query Match 99.8%; Score 3346; DB 14; Length 655;
Best Local Similarity 99.8%; Pred. No. 7.5e-288;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MSSNVEVEFIPVSOQNTNGFPATASNDLKAFTGAVLSFHNI CYRVKLSKSGFLPCRKPYE 60
Db      1 MSSNVEVEFIPVSOQNTNGFPATASNDLKAFTGAVLSFHNI CYRVKLSKSGFLPCRKPYE 60
Qy      61 KEILSININGIMKPGNALIIGPTGGKSSLLDVLAAKDPBSGLSGDVLINGAPRPAKFKN 120
Db      61 KEILSININGIMKPGNALIIGPTGGKSSLLDVLAAKDPBSGLSGDVLINGAPRPAKFKN 120
Qy      121 SGVVODDVVNGTLTVRENLOFSALRLATTTMTHNEKNRINRVIOELGLDYADSKVGT 180
Db      121 SGVVODDVVNGTLTVRENLOFSALRLATTTMTHNEKNRINRVIOELGLDYADSKVGT 180
Qy      181 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLLKRMKSOGRTIIF 240
Db      181 QTRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLLKRMKSOGRTIIF 240
Qy      241 SHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADPFLDIING 300
Db      241 SHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADPFLDIING 300
Qy      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSPFYKETAELHOLSGGEKKKK 360
Db      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSPFYKETAELHOLSGGEKKKK 360
Qy      361 ITVPKEISYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIITVVLGLVIGALYFGLKXDS 420
Db      361 ITVPKEISYTTSFCHOLRWVSKRSFKNLLGNPOASIAQIITVVLGLVIGALYFGLKXDS 420
Qy      421 TGIORAGVLPFLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLLSDLLP 480
Db      421 TGIORAGVLPFLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLLSDLLP 480
Qy      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Db      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGSVSVATLL 540
Qy      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Db      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Qy      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 655
Db      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKXYS 655
```

```
RESULT 5
US-10-405-806-2
; Sequence 2, Application US/10405806
; Publication No. US2003023262A1
; GENERAL INFORMATION:
; APPLICANT: KOMATANI, HIDEYA
; APPLICANT: HARA, YOSHITAKU
; APPLICANT: KOTANI, HIDEHIITO
; APPLICANT: NAKAGAWA, RINKO
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF
; FILE REFERENCE: 234985USOCONT
; CURRENT APPLICATION NUMBER: US/10/405,806
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: PCT/JP01/08112
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP2000-303441
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-405-806-2
```

Query Match 99.8%; Score 3346; DB 15; Length 655;
Best Local Similarity 99.8%; Pred. No. 7.5e-288;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY      1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVAKSGFLPCRKPYE 60
      1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVAKSGFLPCRKPYE 60
Db      1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVAKSGFLPCRKPYE 60
QY      61 KEILSNINGIMKPGLNAILGPTGGKSSLLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
      61 KEILSNINGIMKPGLNAILGPTGGKSSLLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
Db      61 KEILSNINGIMKPGLNAILGPTGGKSSLLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
QY      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
Db      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
QY      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
Db      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
QY      241 SIHQPRYSIFKLPDSLTLTLLASGRLMFHGPAQOALGFESAGYHCEAYNNPADFFLDIING 300
      241 SIHQPRYSIFKLPDSLTLTLLASGRLMFHGPAQOALGFESAGYHCEAYNNPADFFLDIING 300
Db      241 SIHQPRYSIFKLPDSLTLTLLASGRLMFHGPAQOALGFESAGYHCEAYNNPADFFLDIING 300
QY      301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAETVNSSFYKETAELHQLSGGEKXK 360
      301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAETVNSSFYKETAELHQLSGGEKXK 360
Db      301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAETVNSSFYKETAELHQLSGGEKXK 360
QY      361 ITVFEKISYTTSPFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKKNDS 420
      361 ITVFEKISYTTSPFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKKNDS 420
Db      361 ITVFEKISYTTSPFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKKNDS 420
QY      421 TGIQNRAGVLFLLTNQCFSSVSAVELFVEKKLFHIEYISGYRVSSYFLGKLLSDLLP 480
      421 TGIQNRAGVLFLLTNQCFSSVSAVELFVEKKLFHIEYISGYRVSSYFLGKLLSDLLP 480
Db      421 TGIQNRAGVLFLLTNQCFSSVSAVELFVEKKLFHIEYISGYRVSSYFLGKLLSDLLP 480
QY      481 MTMLPSIIFTCIYFMYGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVSVATLL 540
      481 MTMLPSIIFTCIYFMYGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVSVATLL 540
Db      481 MTMLPSIIFTCIYFMYGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVSVATLL 540
QY      541 MTICVFPMIFSGLVNLTITIASWLSWLOYFSIPRYGFTALQHNELGQNFCEGLNATGN 600
      541 MTICVFPMIFSGLVNLTITIASWLSWLOYFSIPRYGFTALQHNELGQNFCEGLNATGN 600
Db      541 MTICVFPMIFSGLVNLTITIASWLSWLOYFSIPRYGFTALQHNELGQNFCEGLNATGN 600
QY      601 NPCNVAATCGEEXLVKQIDLSPWGLMKXHVLAACMIVIFLTAIYKLLFLKXYS 655
      601 NPCNVAATCGEEXLVKQIDLSPWGLMKXHVLAACMIVIFLTAIYKLLFLKXYS 655
Db      601 NPCNVAATCGEEXLVKQIDLSPWGLMKXHVLAACMIVIFLTAIYKLLFLKXYS 655

RESULT 6
US-10-874-706-24
; Sequence 24, Application US/10874706
; Publication No. US20050048610A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LAM, Preeti
; APPLICANT: YANG, Junming
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: TANG, Y. Tom
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yaida
; APPLICANT: LU, Dying Anna M.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: PATTERSON, Chandra
; TITLE OF INVENTION: HUMAN TRANSPORT PROTEINS
; FILE REFERENCE: PF-0709 PCT
; CURRENT APPLICATION NUMBER: US/10/874,706
; FILE REFERENCE: 2004-06-24
; PRIOR APPLICATION NUMBER: US/10/009,328
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/139,923; 60/148,177; 60/149,357; 60/162,287
; PRIOR FILING DATE: 1999-06-17; 1999-08-10; 1999-08-18; 1999-10-28
; NUMBER OF SEQ ID NOS: 86
```

```
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 655
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 5517972CD1
US-10-874-706-24

Query Match      99.8%; Score 3346; DB 17; Length 655;
Best Local Similarity 99.8%; Pred. No. 7,5e-288;
Matches 654; Conservative 0; Mismatch 1; Indels 0; Gaps 0;

QY      1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVAKSGFLPCRKPYE 60
      1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVAKSGFLPCRKPYE 60
Db      1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVAKSGFLPCRKPYE 60
QY      61 KEILSNINGIMKPGLNAILGPTGGKSSLLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
      61 KEILSNINGIMKPGLNAILGPTGGKSSLLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
Db      61 KEILSNINGIMKPGLNAILGPTGGKSSLLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
QY      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
Db      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
QY      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
Db      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
QY      241 SIHQPRYSIFKLPDSLTLTLLASGRLMFHGPAQOALGFESAGYHCEAYNNPADFFLDIING 300
      241 SIHQPRYSIFKLPDSLTLTLLASGRLMFHGPAQOALGFESAGYHCEAYNNPADFFLDIING 300
Db      241 SIHQPRYSIFKLPDSLTLTLLASGRLMFHGPAQOALGFESAGYHCEAYNNPADFFLDIING 300
QY      301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAETVNSSFYKETAELHQLSGGEKXK 360
      301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAETVNSSFYKETAELHQLSGGEKXK 360
Db      301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAETVNSSFYKETAELHQLSGGEKXK 360
QY      361 ITVFEKISYTTSPFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKKNDS 420
      361 ITVFEKISYTTSPFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKKNDS 420
Db      361 ITVFEKISYTTSPFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKKNDS 420
QY      421 TGIQNRAGVLFLLTNQCFSSVSAVELFVEKKLFHIEYISGYRVSSYFLGKLLSDLLP 480
      421 TGIQNRAGVLFLLTNQCFSSVSAVELFVEKKLFHIEYISGYRVSSYFLGKLLSDLLP 480
Db      421 TGIQNRAGVLFLLTNQCFSSVSAVELFVEKKLFHIEYISGYRVSSYFLGKLLSDLLP 480
QY      481 MTMLPSIIFTCIYFMYGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVSVATLL 540
      481 MTMLPSIIFTCIYFMYGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVSVATLL 540
Db      481 MTMLPSIIFTCIYFMYGLKPKADAFVMMFTLMVAVYSSSMALAIAGQSVSVATLL 540
QY      541 MTICVFPMIFSGLVNLTITIASWLSWLOYFSIPRYGFTALQHNELGQNFCEGLNATGN 600
      541 MTICVFPMIFSGLVNLTITIASWLSWLOYFSIPRYGFTALQHNELGQNFCEGLNATGN 600
Db      541 MTICVFPMIFSGLVNLTITIASWLSWLOYFSIPRYGFTALQHNELGQNFCEGLNATGN 600
QY      601 NPCNVAATCGEEXLVKQIDLSPWGLMKXHVLAACMIVIFLTAIYKLLFLKXYS 655
      601 NPCNVAATCGEEXLVKQIDLSPWGLMKXHVLAACMIVIFLTAIYKLLFLKXYS 655
Db      601 NPCNVAATCGEEXLVKQIDLSPWGLMKXHVLAACMIVIFLTAIYKLLFLKXYS 655

RESULT 7
US-09-866-866A-27
; Sequence 27, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
```

/ PRIOR FILING DATE: 1999-05-27
/ PRIOR APPLICATION NUMBER: 60/086,988
/ PRIOR FILING DATE: 1998-05-28
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 27
/ LENGTH: 655
/ TYPE: PR
/ ORGANISM: Homo sapien
US-09-866-866A-27

Query Match 99.6%; Score 3338; DB 9; Length 655;
Best Local Similarity 99.7%; Pred. No. 3.8e-287;
Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPYE 60
DB 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPYE 60
QY 61 KEILSINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
DB 61 KEILSINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
QY 121 SGVYVDDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKVADSKVGT 180
DB 121 SGVYVDDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIMGELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITF 240
DB 181 QFIRGVSGGERKRTSIMGELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITF 240
QY 241 SIHQPRYSIFKLPDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
DB 241 SIHQPRYSIFKLPDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
QY 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAETVNSSPYKETAELHQSAGEKKK 360
DB 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAETVNSSPYKETAELHQSAGEKKK 360
QY 361 ITVFKESITTSFCHQLRWKRSRSPKULLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
DB 361 ITVFKESITTSFCHQLRWKRSRSPKULLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
QY 421 TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEHISGYRVSFFLGKLSDDL 480
DB 421 TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEHISGYRVSFFLGKLSDDL 480
QY 481 MTMLPSIIFTCIVYFMGLKPKADAFVVMFTLMVAVASASSMALAIAAGOSVSVATLL 540
DB 481 MTMLPSIIFTCIVYFMGLKPKADAFVVMFTLMVAVASASSMALAIAAGOSVSVATLL 540
QY 541 MTICFVPMWIFSGLLVNLTTIASWLSWLOFSIPRYGFTALQHNELGONFCGLNATGN 600
DB 541 MTICFVPMWIFSGLLVNLTTIASWLSWLOFSIPRYGFTALQHNELGONFCGLNATGN 600
QY 601 NPCNVAATCTGGEYLVKQGIIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
DB 601 NPCNVAATCTGGEYLVKQGIIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
```

RESULT 8
US-09-866-866A-10

/ Sequence 10, Application US/09866866A
/ Patent No. US20020102244A1
/ GENERAL INFORMATION:
/ APPLICANT: Sorrentino, Brian
/ APPLICANT: Schuetz, John
/ TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
/ FILE REFERENCE: 1340-1-02CIP2
/ CURRENT APPLICATION NUMBER: US/09/866,866A
/ PRIOR FILING DATE: 2001-08-30
/ PRIOR APPLICATION NUMBER: 09/584,586
/ PRIOR FILING DATE: 2000-05-31

/ PRIOR APPLICATION NUMBER: PCT/US99/11825
/ PRIOR FILING DATE: 1999-05-27
/ PRIOR APPLICATION NUMBER: 60/086,988
/ PRIOR FILING DATE: 1998-05-28
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 10
/ LENGTH: 655
/ TYPE: PR
/ ORGANISM: Homo sapien
US-09-866-866A-10

Query Match 99.4%; Score 3331; DB 9; Length 655;
Best Local Similarity 99.4%; Pred. No. 1.6e-286;
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPYE 60
DB 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPYE 60
QY 61 KEILSINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
DB 61 KEILSINGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
QY 121 SGVYVDDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKVADSKVGT 180
DB 121 SGVYVDDVVMGTLTVRENLQFSALRLATTTNHEKNERINRVIOELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIMGELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITF 240
DB 181 QFIRGVSGGERKRTSIMGELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITF 240
QY 241 SIHQPRYSIFKLPDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
DB 241 SIHQPRYSIFKLPDSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
QY 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAETVNSSPYKETAELHQSAGEKKK 360
DB 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAETVNSSPYKETAELHQSAGEKKK 360
QY 361 ITVFKESITTSFCHQLRWKRSRSPKULLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
DB 361 ITVFKESITTSFCHQLRWKRSRSPKULLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
QY 421 TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEHISGYRVSFFLGKLSDDL 480
DB 421 TGIQRAGVLPFLTNQCFSSVSAVELFVEKKLFHEHISGYRVSFFLGKLSDDL 480
QY 481 MTMLPSIIFTCIVYFMGLKPKADAFVVMFTLMVAVASASSMALAIAAGOSVSVATLL 540
DB 481 MTMLPSIIFTCIVYFMGLKPKADAFVVMFTLMVAVASASSMALAIAAGOSVSVATLL 540
QY 541 MTICFVPMWIFSGLLVNLTTIASWLSWLOFSIPRYGFTALQHNELGONFCGLNATGN 600
DB 541 MTICFVPMWIFSGLLVNLTTIASWLSWLOFSIPRYGFTALQHNELGONFCGLNATGN 600
QY 601 NPCNVAATCTGGEYLVKQGIIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
DB 601 NPCNVAATCTGGEYLVKQGIIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
```

RESULT 9

US-10-090-455-5
/ Sequence 5, Application US/10090455
/ Publication No. US20030027259A1
/ GENERAL INFORMATION:
/ APPLICANT: Chen, Hongyun
/ APPLICANT: Le Bihan, Stephane
/ TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
/ FILE REFERENCE: 100103.406
/ CURRENT APPLICATION NUMBER: US/10/090,455
/ PRIOR FILING DATE: 2002-03-01
/ NUMBER OF SEQ ID NOS: 17

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 655
TYPE: PRF
ORGANISM: Homo sapiens
US-10-090-455-5

Query Match 99.4%; Score 3331; DB 14; Length 655;
Best Local Similarity 99.4%; Pred. No. 1,6e-286;
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MSSSNEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCKRPVE 60
DB 1 MSSSNEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCKRPVE 60
QY 61 KETLSNINGIMKGNLAIIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPNPKCN 120
DB 61 KETLSNINGIMKGNLAIIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPNPKCN 120
QY 121 SGVYVDDVVMGTLTVRENILQFSALRLATMTNHEKNERINRVIOELGLDKVADSRYGT 180
DB 121 SGVYVDDVVMGTLTVRENILQFSALRLATMTNHEKNERINRVIOELGLDKVADSRYGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRTIIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRTIIF 240
QY 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
DB 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
QY 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGEKKKK 360
DB 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGEKKKK 360
QY 361 ITVFKESITTSFCHQLRWVSKSPKOLLGNPOASIAQIIVTVVLGIVGAIYFGLKND 420
DB 361 ITVFKESITTSFCHQLRWVSKSPKOLLGNPOASIAQIIVTVVLGIVGAIYFGLKND 420
QY 421 TGIQNRAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGLSLDLP 480
DB 421 TGIQNRAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGLSLDLP 480
QY 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYSASSMALAIAGQSVSVATLL 540
DB 481 MRMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYSASSMALAIAGQSVSVATLL 540
QY 541 MTICFVFMIFSGILVNLTTIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
DB 541 MTICFVFMIFSGILVNLTTIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
QY 601 NPCNVATCTGEEVLVKGIDLSPMGLMKHVALACMIVIFLITAYIKLFLKXYS 655
DB 601 NPCNVATCTGEEVLVKGIDLSPMGLMKHVALACMIVIFLITAYIKLFLKXYS 655
```

RESULT 10

US-09-745-763-197

Sequence 197, Application US/09745763

Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

Lavalie, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 197:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 197:
US-09-745-763-197

Query Match 91.1%; Score 3053.5; DB 9; Length 604;
Best Local Similarity 91.9%; Pred. No. 6,3e-262;
Matches 602; Conservative 0; Mismatches 2; Indels 51; Gaps 1;

```
QY 1 MSSSNEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCKRPVE 60
DB 1 MSSSNEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSGFLPCKRPVE 60
QY 61 KETLSNINGIMKGNLAIIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPNPKCN 120
DB 61 KETLSNINGIMKGNLAIIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPNPKCN 120
QY 121 SGVYVDDVVMGTLTVRENILQFSALRLATMTNHEKNERINRVIOELGLDKVADSRYGT 180
DB 121 SGVYVDDVVMGTLTVRENILQFSALRLATMTNHEKNERINRVIOELGLDKVADSRYGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRTIIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQGRTIIF 240
QY 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
DB 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
QY 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGEKKKK 360
DB 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGEKKKK 360
QY 361 ITVFKESITTSFCHQLRWVSKSPKOLLGNPOASIAQIIVTVVLGIVGAIYFGLKND 420
DB 361 ITVFKESITTSFCHQLRWVSKSPKOLLGNPOASIAQIIVTVVLGIVGAIYFGLKND 420
QY 421 TGIQNRAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGLSLDLP 480
DB 421 TGIQNRAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGLSLDLP 480
QY 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYSASSMALAIAGQSVSVATLL 540
DB 481 MRMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYSASSMALAIAGQSVSVATLL 540
QY 541 MTICFVFMIFSGILVNLTTIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
DB 541 MTICFVFMIFSGILVNLTTIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
```

Db 490 MTICFVPMIFSGILVNLTTIASWLSWLOFYSIPRYGFTALQNEFLQNFPCGLNATGN 549
QY 601 NNCNATCTGGERLYKQIDLSPWGIMKRNHVALACMIVIFLTAIVKLLFLKKYS 655
Db 550 NNCNATCTGGERLYKQIDLSPWGIMKRNHVALACMIVIFLTAIVKLLFLKKYS 604

RESULT 11

US-09-866-866A-14
/ Sequence 14, Application US/09866666A
/ Patent No. US20020102244A1
/ GENERAL INFORMATION:
/ APPLICANT: Sorrentino, Brian
/ APPLICANT: Schuetz, John
/ TITLE OF INVENTION: A method of identifying and/or isolating stem cells
/ FILE REFERENCE: 1340-1-021CIP2
/ CURRENT APPLICATION NUMBER: US/09/866,866A
/ CURRENT FILING DATE: 2001-08-30
/ PRIOR APPLICATION NUMBER: 09/584,586
/ PRIOR FILING DATE: 2000-05-31
/ PRIOR APPLICATION NUMBER: PCT/US99/11825
/ PRIOR FILING DATE: 1999-05-27
/ PRIOR APPLICATION NUMBER: 60/086,988
/ PRIOR FILING DATE: 1998-05-28
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 14
/ LENGTH: 657
/ TYPE: PRF
/ ORGANISM: Mus musculus
US-09-866-866A-14

Query Match 82.2%; Score 2757; DB 9; Length 657;
Best Local Similarity 81.5%; Pred. No. 1,6e-235;
Matches 536; Conservative 51; Mismatches 67; Indels 4; Gaps 3;

QY 1 MSSNVEVEPIPVQSGNTGFPATASNDLKAFTGAVLSFPHNICYRVKLSGFLPCCKPYE 60
Db 1 MSSNDHVLVPMQSRNNGLPRMNSKAVTLABGDVLSFHHITRYKVASGFL-VAKTVE 59
QY 61 KEILSNINGIMKPGALNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGARPAFKN 120
Db 60 KEILSDINGIMKPGALNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGARPAFKN 119
QY 121 SGYVVDVDDVMGTLTVRENLOFSAALRLATTTMNEKNERINRVIOELGLDKVADSKVT 180
Db 120 SGYVVDVDDVMGTLTVRENLOFSAALRLPTTMNEKNERINTIIEKLEKVDASKVT 179
QY 181 QPIRVSQGERKRTSISGMELITDPSILFLDEPTTGIDSTANAVLLLLKRMKQRTITF 240
Db 180 QPIRISIGGERKRTSISGMELITDPSILFLDEPTTGIDSTANAVLLLLKRMKQRTITF 239
QY 241 SIHOPRYSIFKLFDSTLLASGRLMFHPGPAQALGFYSAGYHCEAYNNPADFFLDI 300
Db 240 SIHOPRYSIFKLFDSTLLASGRLMFHPGPAQALGFYSAGYHCEAYNNPADFFLDI 299
QY 301 DSTAVANLREE-DFAKATELIEPSKODKPLIEKLAETIYVNSSEYKETAELHQLSGEKKK 359
Db 300 DSAVAVLNREEQDNENKTEEPSKQKPYIENLSEFYNSAIYGETKALDQLPGAQKK 359
QY 360 KITVKEISYTTSPFCHOLRWVSKSPKLNLGNPQASIAQIIVTVVGLVIGAIYGLKND 419
Db 360 GTSAREPEYVTSFCHOLRWVSKSPKLNLGNPQASIAQIIVTVVGLVIGAIYGLKND 419
QY 420 STGIQNRAGVLFPLTTNQCFSVSAVELFVVEKKLFIHEIYSGYRVSSYFGLKLSDL 479
Db 420 AAGQNRAGVLFPLTTNQCFSVSAVELFVVEKKLFIHEIYSGYRVSSYFGLKMSDL 479
QY 480 PNTMLPSIITFCIYVFMGLKPKADAFVVMFTLMMVAVSASSMALAIAGQSVSVATL 539
Db 480 PNRFLPSVIFTCIYVFMGLKPKADAFVFMFTLMMVAVTASSMALAIAGQSVSVATL 539
QY 540 LMTICFVPMIFSGILVNLTTIASWLSWLOFYSIPRYGFTALQNEFLQNFPCGLNATG 599

Db 540 LMTIAFVPMIFSGILVNLRTIGPWLQYFYSIPRYGFTALQNEFLQNEFCFGFNATD 599
QY 600 NRPC--NYATCTGGERLYKQIDLSPWGIMKRNHVALACMIVIFLTAIVKLLFLKKYS 655
Db 600 NSTCVNSYACTGNEBYLINOGLIELSPWGLMKRNHVALACMIIIFLTAIVKLLFLKKYS 657

RESULT 12

US-10-369-493-1520
/ Sequence 1520, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianteng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 1520
/ LENGTH: 1049
/ TYPE: PRF
/ ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1520

Query Match 24.9%; Score 835.5; DB 15; Length 1049;
Best Local Similarity 30.5%; Pred. No. 1,6e-64;
Matches 222; Conservative 134; Mismatches 257; Indels 115; Gaps 18;

QY 1 MSSNVEVEPIPVQSGNTGFPATASNDLKAFTGAVLSFPHNICYRVKLSGFLPCCKPYE 60
Db 355 LQSSNSPIRLP-DEAVANNFLONEDDTL-----ATLSENTITYSVPSINS-----DGYE 402
QY 61 KEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGARPAFKN-ANPK 118
Db 403 ETVLNEISIGIVAPGGLIIMGSSGAKTLLDILAMKRTGHSVSGIKXNGISMBRKSS 462
QY 119 CNSGYVVDVDDVMGTLTVRENLOFSAALRLATTTMNEKNERINRVIOELGLDKVADSKV 178
Db 463 KIIGVDDDDFLPTLVFEYVJALNLRPLPVALFEAKKAVYVLEBRLIIDIKRIT 522
QY 179 GTPQIRVSQGERKRTSISGMELITDPSILFLDEPTTGIDSTANAVLLLLKRMKQ-GR 237
Db 523 GNEFDRGISGGKRVSTIACELVTSPLVFLDEPTSGLDASANNVIBCLVRLSSDYNT 582
QY 238 IIFSHPRYSIFKLFDSTLLASGRLMFHPGPAQALGFYSAGYHCEAYNNPADFFLDI 297
Db 583 LVLSTHQRNSNIFLFDKVLVLSKEMVYSGNAKAVSELRNEGICPNNYNIADYLDI 642
QY 238 -----INDSTAV 305
Db 643 TFEAPQGRRRIRNISDLAAGTDNDIDNTIHTFTSSDGTQREMAHLAHRDEIRS 702
QY 366 ALNREDEKATE-----IIEPSKODKPLIEKLAETIYVNSSEYKETAELHQLSGEKKK 360
Db 703 LIRDEDEVDGDRGATEIDINTLHLDK-----KDSYVAELSOEIEVLSGDEBSN 758
QY 361 IT--AFKESISYTTSPFCHOLRWVSKSPKLNLGNPQASIAQIIVTVVGLVIGAIYGLKND 418
Db 759 VLANGLPFGQGSAGLQGLSTLNSRSFNNMVRNPGLLGNVLTLLSLFGLTYNNYSN 818
QY 419 DSTGIQNRAGVLFPLTTNQCFSVSAVELFVVEKKLFIHEIYSGYRVSSYFGLKLSDL 478
Db 819 DISGQNRMGHFFFLTYFGEVTFGLSSFALERIIFPKERSNNYSPPLAYISKIMSEV 878
QY 479 LPMTMLPSIITFCIYVFMGLKPKADAFVVMFTLMMVAVSASSMALAIAGQSVSVAT 538

```

Db      879 VPLRVPRLSLIVPMGTGLNMKNDAFFKICIGLLIFNLGISLEILLITIGIFEDLNNSI 938
Qy      539 LMTICVFMWISGLV---NLTTASWLSWLOYESIPRYGFTALOHNF----- 586
Db      939 ILTVLVLGSLPSGLPINTKNTTVA--FKYLNFSVFYFAAESLITNEVKTLMKERK 996
Qy      587 LQGNF-CPGIATGNPCNATCTGEEYLVKQGI--DLSPWGLMKHVALACMIIVFLTI 643
Db      997 YGNIIEVPG-----ATLSTFGFVQNLVFDIK-----ILALFNVLFLIM 1036
Qy      644 AYTKLFL 651
Db      1037 GYIALKWI 1044

RESULT 13
US-10-108-605-245
; Sequence 245, Application US/10108605
; Publication No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stem, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108, 605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 245
; LENGTH: 663
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
US-10-108-605-245

Query Match      24.2%; Score 812; DB 13; Length 663;
Best Local Similarity 32.1%; Pred. No. 9.5e-63;
Matches 210; Conservative 134; Mismatches 251; Indels 60; Gaps 17;

Qy      5 NVEVFIPVSGQNTNGPATNSDLKAFTEGAVLSFHNICRYVTKSGFLPCRPEVEKIL 64
Db      50 NMDIFGAVNQ-----FGSGWRQLVNRTRGLFCNERHI-----PAPR---KHL 89
Qy      65 SNINGIMKPG-LNAIIGPTGGKSSLLDYLAAKDPGSL---SGDVLINGAPRPA-NFK 118
Db      90 KNYGVAYPEELAVWSSGAGKTTLLNALAFR-SFGIOVSPSGMRLNGQVDAKEMQ 148
Qy      119 CNSGYVODDVVNGTLTVRENLOFSAALRLATTTNHEKNERINRYIOELGLDKVADSKV 178
Db      149 ARCAVYQODDLFGISLTAREHLIFQAMVWRPHLTQRVARVDQVIOELSLSKQHTII 208
Qy      179 GTQ-FIRGVSGGRRKRTSIGMELITDPSILFLDEPTGLDSSSTANAVLLKMKSKGRT 237
Db      209 GVPGRVYGSIGGERKRLAFASEALTDPLLCDEPTSGDLSFPAHSVVOVLKXLSQSGKT 268
Qy      238 IIFSIOPRYSIFKLFDLSLTLLASGRLMFHPGPAQELAGYESAGYHCEAVNPNADFLDI 297
Db      269 VILTIHOPSELELFDKILLMAGRAVAFGTSEAVDFSYGAOCPNTYNNPADFLYOV 328
Qy      298 INGSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSFFYETKAEHLQSGEK 357
Db      329 L-----AVVGRRIEER-----DRIKICNPAISKVAR-DMEQLLATKN 367
Qy      358 KKTIYFKEISY--TSFCHQLRWVSKSFKNLGNPQASIAQIIVTVVLGLVIAGIYFG 415
Db      368 LEKPLQEPENGYYKATFWQFRAVLWRWSLVLEKEPLLVKVRLIQTTVAIILIGLIFG 427
```

```

Qy      416 LKNDSTGIONRAGVLEFLTNNOCFSSVSA-VELFVEKELFIHEIYISGYRVSSYFLGKL 474
Db      428 QQLTVGWNMINCAIILPLTNMFGQVNPATINVFTESELFPVMEARSRLRCOTYFLGKT 487
Qy      475 LSDLMTMTLPSIIFPCICYFMGLKPKADAPVMMFTLMVAVYSASSMALAIQAOSV 534
Db      488 IAE-LPLFLTVPLVFAIAYPMIGLRAGVLAHFENCALATVLVANVSTSFQYLISCASSST 546
Qy      535 SVATLMTICFVPMIIFSGLVNLTTIASWLSWLOYESIPRYGFTALOHNEFLGQNFPCG 594
Db      547 SMLASVGPVITIEFLFPGCFELNSGSVPYTLKULSTLSMFRYANEGILLNQWADVE--RG 604
Qy      595 -LNATGNPCNATCTGEEYLVKQGI--DLSPWGLMKHVALACMIIVFLTIAYKL 648
Db      605 EISCTSSN-----TTCSSGKVLLETINFSAADPLDYGVIALIVSFRVLIATL 655

RESULT 14
US-10-732-923-1785
; Sequence 1785, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgercon, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 1785
; LENGTH: 687
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
US-10-732-923-1785

Query Match      24.2%; Score 812; DB 17; Length 687;
Best Local Similarity 32.1%; Pred. No. 1e-62;
Matches 210; Conservative 134; Mismatches 251; Indels 60; Gaps 17;

Qy      5 NVEVFIPVSGQNTNGPATNSDLKAFTEGAVLSFHNICRYVTKSGFLPCRPEVEKIL 64
Db      74 NMDIFGAVNQ-----FGSGWRQLVNRTRGLFCNERHI-----PAPR---KHL 113
Qy      65 SNINGIMKPG-LNAIIGPTGGKSSLLDYLAAKDPGSL---SGDVLINGAPRPA-NFK 118
Db      114 KNYGVAYPEELAVWSSGAGKTTLLNALAFR-SFGIOVSPSGMRLNGQVDAKEMQ 172
Qy      119 CNSGYVODDVVNGTLTVRENLOFSAALRLATTTNHEKNERINRYIOELGLDKVADSKV 178
Db      173 ARCAVYQODDLFGISLTAREHLIFQAMVWRPHLTQRVARVDQVIOELSLSKQHTII 232
Qy      179 GTQ-FIRGVSGGRRKRTSIGMELITDPSILFLDEPTGLDSSSTANAVLLKMKSKGRT 237
Db      233 GVPGRVYGSIGGERKRLAFASEALTDPLLCDEPTSGDLSFPAHSVVOVLKXLSQSGKT 292
Qy      238 IIFSIOPRYSIFKLFDLSLTLLASGRLMFHPGPAQELAGYESAGYHCEAVNPNADFLDI 297
Db      293 VILTIHOPSELELFDKILLMAGRAVAFGTSEAVDFSYGAOCPNTYNNPADFLYOV 352
Qy      298 INGSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSFFYETKAEHLQSGEK 357
Db      353 L-----AVVGRRIEER-----DRIKICNPAISKVAR-DMEQLLATKN 391
Qy      358 KKTIYFKEISY--TSFCHQLRWVSKSFKNLGNPQASIAQIIVTVVLGLVIAGIYFG 415
Db      392 LEKPLQEPENGYYKATFWQFRAVLWRWSLVLEKEPLLVKVRLIQTTVAIILIGLIFG 451
Qy      416 LKNDSTGIONRAGVLEFLTNNOCFSSVSA-VELFVEKELFIHEIYISGYRVSSYFLGKL 474
Db      452 QQLTVGWNMINCAIILPLTNMFGQVNPATINVFTESELFPVMEARSRLRCOTYFLGKT 511
```

Qy 475 LSDLEMTMLPSIIFTCIYFMLGLPKADAFVMMFTLMMVAYSSMALAIAGOSVY 534
Db 512 IAE-LPLFLTVPLVFAIAYPMIGLRAGVLHFNCLALVTLVANVSTFGLISCSST 570
Qy 535 SVATLMTICFVPMIIFSGLLVNLTTIASMLWLYFSIIPRYGFTALQHNEFLGONFCG 594
Db 571 SMALSVGPVLIIPFLIFGFFLNSGSVPYVLYKWLSTLSWFRYANEGLLINQMDVBE--Pg 628
Qy 595 -LNATGNMPCNATCTGEEYLVKQIDLSPMGLMKKHVALACIYVFLITAYKL 648
Db 629 EISCTSSN---TTCPSGKVIETLTFSAADLPDYVGLAILIVSFRVLAIAL 679

RESULT 15
US-10-732-923-1786

/ Sequence 1786, Application US/10732923
/ Publication No. US20050108791A1
/ GENERAL INFORMATION:
/ APPLICANT: Edgerton, Michael D
/ TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
/ FILE REFERENCE: 38-15(52796)C
/ CURRENT APPLICATION NUMBER: US/10/732,923
/ CURRENT FILING DATE: 2003-12-10
/ PRIOR APPLICATION NUMBER: 10/310,154
/ PRIOR FILING DATE: 2002-12-04
/ NUMBER OF SEQ ID NOS: 24149
/ SEQ ID NO 1786
/ LENGTH: 687
/ TYPE: PRF
/ ORGANISM: Drosophila melanogaster
US-10-732-923-1786

Query Match 24.2%; Score 812; DB 17; Length 687;

Best Local Similarity 32.1%; Pred. No. 1e-62; Matches 210; Conservative 134; Mismatches 251; Indels 60; Gaps 17;

Qy 5 NVEVFIPVSGNTNGFPATASNDLKAFTEGAVLSFNHICRYVLYKSGFLPCRKYVEKITL 64
Db 74 NMDIFGANVQ-----PGSGMRQLVNRTRGLFCNERHI-----PAPR--KHLL 113
Qy 65 SNINGIMKPG-LNAILGPTGGKSSLDVLAARKDPSGL---SGDVLINGAPRPA-NPK 118
Db 114 KAVCGVAYGELLAVWSSGAGKTTLLNALFR-SFGIIVSPSGMRLNGQPVDAKEMQ 172
Qy 119 CNSGYVQDDVVMGTITVRENLQFSALRLATTMTNHEKERNRYIQELGDKVADSKV 178
Db 173 ARCAVVOODDLFGSLTAREHLIFQAMVMRPHLTFRQVAVVDQVIOELSLSKQHTII 232
Qy 179 GNG-FIRGVSGGERKRTSGMELITPSTLFLDEPTTGDSSTANAVLLLRMSKQGR 237
Db 233 GVPGRVKGISGGERKRLAPASEALTDPLICDEPTSGDSFTAHSVQVLLKLSQKGT 292
Qy 238 IIFSIIHOPRYSIFKLFDSITLLASGRMFHGPQOALGYFESAGYCEAYNNPADFFLDI 297
Db 293 VILTIHOPSESELFELFDKILMAEGRVAFGLTPEBAVDFFSYGAQCPTNNVPADFYQV 352
Qy 298 INGDSTAYALNREEDPKATEITEPSKQDPLIEKLAETVNSSFYKETABEIHOLSGEK 357
Db 353 L-----AVVPGREIESR-----DRIKICDNFAISKVAR-DMEQLATKN 391
Qy 358 KKKITVFKRISYT--TSFCHOLRWVSKRSPKULLGNPQASIAQIITVVVLGVIGAIYRG 415
Db 392 LKXPLEQEPENGYTYKATFMQPRAYLVMRSLSVLKEPLVYKRLIQTINVAITLIGITFLG 451
Qy 416 LKNDSTGIONRAGVLFLLTNOCFSSVSA-VELFVVEKELFTHETISGYRVSSYFLGKL 474
Db 452 QQLTQGVANNINGAIFLPLTNMTFQVVFATINVFSELVFMREARSRLYRCDTYFLGKT 511
Qy 475 LSDLEMTMLPSIIFTCIYFMLGLPKADAFVMMFTLMMVAYSSMALAIAGOSVY 534
Db 512 IAE-LPLFLTVPLVFAIAYPMIGLRAGVLHFNCLALVTLVANVSTFGLISCSST 570
Qy 535 SVATLMTICFVPMIIFSGLLVNLTTIASMLWLYFSIIPRYGFTALQHNEFLGONFCG 594

Db 571 SMALSVGPVLIIPFLIFGFFLNSGSVPYVLYKWLSTLSWFRYANEGLLINQMDVBE--Pg 628
Qy 595 -LNATGNMPCNATCTGEEYLVKQIDLSPMGLMKKHVALACIYVFLITAYKL 648
Db 629 EISCTSSN---TTCPSGKVIETLTFSAADLPDYVGLAILIVSFRVLAIAL 679

Search completed: June 6, 2005, 09:57:41
Job time: 144 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2005, 09:27:54 ; Search time 44 Seconds

(without alignments)
1432.318 Million cell updates/sec

Title: US-09-961-086-1

Perfect score: 3352
Sequence: 1 MSSNVEVPIVPSQGNNGF.....MIVIFLTIVLKLFLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR 79:*

1: p1r1:*\n2: p1r2:*\n3: p1r3:*\n4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2849.5	85.0	656	2 UC7860	brain multidrug re
2	835.5	24.9	1049	1 S19421	ATP-dependent perm
3	812	24.2	687	1 FYFWM	white protein - fr
4	800.5	23.9	737	2 T46101	ABC transporter-11
5	774	23.1	646	2 C86441	probable ABC trans
6	767.5	22.9	687	2 D96553	hypothetical prote
7	746	22.3	649	2 A84509	probable ABC trans
8	745	22.2	725	2 T47652	ABC transporter-11
9	741.5	22.1	739	2 T45891	ABC transporter-11
10	739	22.0	678	2 H96552	hypothetical prote
11	726.5	21.7	708	2 T47650	ABC transporter-11
12	721	21.5	635	2 T08934	hypothetical prote
13	720.5	21.5	755	2 G84791	probable ABC trans
14	717.5	21.4	740	1 T0567	probable ATP-bindi
15	712.5	21.3	609	2 E96742	probable ABC trans
16	710	21.2	638	2 G02068	white homolog - hu
17	708	21.1	646	2 JC7777	ATP binding caset
18	707.5	21.1	547	2 T31543	hypothetical prote
19	707.5	21.1	720	2 T47648	ABC transporter-11
20	677	20.2	559	2 B88474	protein C05D10.3 f
21	676	20.2	725	2 C84423	probable ABC trans
22	669	20.0	659	2 E86313	hypothetical prote
23	664	19.8	608	2 T34391	hypothetical prote
24	660	19.7	662	2 T47649	ABC transporter-11
25	658.5	19.6	1294	2 S77690	probable membrane
26	657.5	19.6	1450	2 T45888	ABC transporter-11
27	656.5	19.6	590	2 B96573	protein F12M16.17
28	649.5	19.4	633	2 T19189	hypothetical prote
29	645.5	19.3	577	2 T04229	ABC-type transp

30	639	19.1	658	2 T31958	hypothetical prote
31	636.5	19.0	639	2 G88839	protein C10C6.5 f1
32	636.5	19.0	695	2 T21109	hypothetical prote
33	627.5	18.7	610	2 T19333	hypothetical prote
34	612.5	18.3	1501	2 S50992	SMO2 protein - yea
35	603	18.3	1530	2 S52239	bretefelin a resist
36	612	18.0	1531	2 T52010	hypothetical prote
37	602	18.0	1511	2 A53151	pleiotropic drug r
38	594	17.7	1564	2 S55517	probable transp
39	593	17.7	1443	2 T02491	probable ABC trans
40	590	17.6	1469	2 H96622	probable ABC trans
41	585	17.5	1333	2 S63403	probable membrane
42	580.5	17.3	1420	2 T02644	ABC-type transp
43	580.5	17.3	1529	2 S69688	hypothetical prote
44	572	17.1	1413	2 G84790	probable ABC trans
45	564	16.8	1466	2 T30566	ATP-binding caset

ALIGNMENTS

RESULT 1

UC7860 brain multidrug resistance protein, BMDP - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 18-Nov-2002 #sequence_revision 18-Nov-2002 #ext_change 09-Jul-2004
C:Accession: UC7860

R:Bisbenzylolactone, T.; Galia, H.J.

Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002

A:Title: A new multidrug resistance protein at the blood-brain barrier.

A:Reference number: UC7860; PMID:12054514

A:Accession: UC7860

A:Molecule type: mRNA

A:Residues: 1-656 <EIS>

A:Cross-references: UNIPROT:Q8MIB3; GB:AJ420927

A:Experimental source: brain

A:Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) super-

family, is involved in the exclusion of xenobiotics from the brain and participates in drug transport across the b

A:Gene: bmdp

Query Match	Score	85.0%	Score	2849.5	DB 2	Length	656
Best Local Similarity	84.3%	Pred. No.	6.9e-192				
Matches	553	Conservative	44	Mismatches	58	Indels	1
QY	1	MSSNVEVPIVPSQGNNGF	PATASNDLKAFTGAVLSFHNI	CYRVKLSGFL	PCRKYV	60	
DB	1	MSSNVEVPIVPSQGNNGF	PATASNDLKAFTGAVLSFHNI	CYRVKLSGFL	PCRKYV	60	
QY	61	KEILSNINGIMKPGNALI	IGPTGGKSSLDVLAARKPSG	SLDVTIN	GAARPANFKCN	120	
DB	61	KEILSNINGIMKPGNALI	IGPTGGKSSLDVLAARKPSG	SLDVTIN	GAARPANFKCN	120	
QY	121	SGYVQDDVMKGLTVRENL	QPSAALRLATTTNHEKNER	IRRVIOELGLDKVADSKVCT	180		
DB	121	SGYVQDDVMKGLTVRENL	QPSAALRLATTTNHEKNER	IRRVIOELGLDKVADSKVCT	180		
QY	181	QPIRVSGGERKRTSIGM	ELTDPISILFLDEFTTGDS	TANAVILLLRMSKQGRITIF	240		
DB	181	QPIRVSGGERKRTSIGM	ELTDPISILFLDEFTTGDS	TANAVILLLRMSKQGRITIF	240		
QY	241	SIHQPRYSIFKFLFSDL	TLTLLASGRMLFHGPARAL	GYFASIGVNCPEYNNPAPFLDIVNG	300		
DB	241	SIHQPRYSIFKFLFSDL	TLTLLASGRMLFHGPARAL	GYFASIGVNCPEYNNPAPFLDIVNG	300		
QY	301	DSTAVALNR-EDFPAT	ELIIEPSKODKPLIEKLA	IYVNSPFYKTKAEHLQSGEKKK	359		
DB	301	DSTAVALNR-EDFPAT	ELIIEPSKODKPLIEKLA	IYVNSPFYKTKAEHLQSGEKKK	359		
QY	360	KITVEKEISYTTSTF	CHQLRWVSKRSFKULL	GNPQASIAQIIVTVVGLVGAIVG	419		
DB	360	KITVEKEISYTTSTF	CHQLRWVSKRSFKULL	GNPQASIAQIIVTVVGLVGAIVG	419		
QY	361	KSSVYKEVYTTSTF	CHQLRWVSKRSFKULL	GNPQASIAQIIVTVVGLVGAIVG	419		
DB	361	KSSVYKEVYTTSTF	CHQLRWVSKRSFKULL	GNPQASIAQIIVTVVGLVGAIVG	419		

A:Experimental source: strain Canton S
R:O'Hare, K.
submitted to the EMBL Data Library, June 1985
A:Reference number: S10240
A:Accession: S10240
A:Molecule type: DNA
A:Residues: 1-24, 'LIFEIPIYHCRTVAD', 30-687 <HA2>
A:Cross-references: EMBL:X02974; NID:G10873; PIDN:CAA26716.1; PID:G10874
A:Experimental source: strain Canton S
C:Genetics:
A:Gene: white; w
A:Cross-references: FlyBase:FBgn000396
A:Amino acids: 24/3; 116/1, 334/2, 439/3; 463/3
C:Superfamily: fruit fly white protein; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
F:113-317/Domain: ATP-binding cassette homology <ABC>
F:1251-137/Region: nucleotide-binding motif A (P-loop)
F:1251-265/Region: nucleotide-binding motif B
F:67, 93, 472, 554, 651/Binding site: carbonate (asn) (covalent) #status predicted

Query Match	Similarity	24.2%	Score 81.2	DB 1	Length 667
Best Local	Stimilarity	23.1% <td>Pred. No.5.6e-49</td> <td></td> <td></td>	Pred. No.5.6e-49		
Matches	Conservative	134	Mismatches 251	Indels 60	Gaps 17
Qy	5	NVEVEFIVPSQNGTNGFPATASNDLKAFTGEGAVLSFPHNI CYRVKLSGSLPKRKPEVEKIL	64		
Db	74	NMDIFGAVNQ-----PSGMRQLVNRTRRG.FPCNERHI-----PAPR---KILL	113		
Qy	65	SNINQIMKPG-LNALGPTGGKSSLLDVLAARDPSL-----SGDVLINGAPRA-NFK	118		
Db	114	KNVCCVAVPGELIWAASSGAGKTTLLNALFR-SPGQIQVSPSMRLLNQVPADAKEMQ	172		
Qy	119	CNSGVVODDVMGTLLTVRENLOFSALRLATTTMHKNERINRVIOELGIDKVASKV	178		
Db	173	ARCAVQGDPLFISGLTAREHILFQAWRMRRHLYRQAVARVDQIDELSLSKQHII	232		
Qy	179	GTQ-FIRVSGGGERKRTISGHELTIDPSIILFIDEPTGLSDSTANAVLLIKMSKQRT	237		
Db	233	GVPGRVKLSGGERKRLFASEBALDPLLLCDSEPTSLGDEFTAHSSVQVKKLSQKCT	292		
Qy	238	IIFSIIHQGRYSIFPKLFDLSLTLLASGRLFHFGPAQDALGFPISAGVHCAYNNPAPDFIDI	297		
Db	293	VILTIHQSSSELFELFDKILLMAERVAFLGTBPSEAVDFEYVQAQCTNNPDPFYQV	352		
Qy	298	INGDSTAVALNREBDFKATEIIEPSKODPLIEKLAIIYVNSFPYKETAKEHLQSGEK	357		
Db	353	L-----AVPPEIRESR-----DRIAKICDPAISKVAR-DMEOQLATKN	391		
Qy	358	KKKITTVPFEISYV--TSRCHQLRWYRSKSPKLLGNPQASIAQIIVVYVGLVIGATYFG	415		
Db	392	LEKLPBQENGYTKATMEMQFRAVLWMSLSTVKEPLLVKRLIQTTMAVILLGLFLG	451		
Qy	416	LKNDSTGIQNRAGVFLFTTNQCCSSVA-VELFVEVEKKLFIHEHISGYYVSSVFLGLK	474		
Db	452	QQLQVQGVMMINGAIFLFLTMNTFQNVPAATINVTSELFPVPMREARSLRCDIYFLGT	511		
Qy	475	LSDLPLMTMLPSIIFTCIVYFMLGKPKRAADFVMMFLTMVAVSASSMALATAAGOSV	534		
Db	512	IAE-LPFLVLPVLFATALAYPMIGIRAGVLFHFCMLALVTLVANVSTSGFGLISCASST	570		
Qy	535	SVATLLMTICVFPMNIFSGLLVNNLTTSIAWSLYQFSIPRYGFTALQHNFFLQNCFG	594		
Db	571	SMALSVGPVAILPFLRPGGFLLNGSVYVYIKMLSYLSWFFYANEGILLINQMAVE--PG	628		
Qy	595	-LNATGNPCVYATCTGEEYVYKQIDISPMGLMKNHVALACMIVFLTAIYKL	648		
Db	629	EISCTSSN-----TTCPSGKGVILETTLNFSADPLDYGLAIIILVSFEVIALAYL	679		

RESULT 4
T46101
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T25B15.80

C.Species: *Arabidopsis thaliana* (mouse-ear cress)
C.Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C.Accession: T46101
R.Alcazar, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.
submitted to the Protein Sequence Database, January 2000
A.Reference number: Z23021

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-737 <ALC>
A:Cross-references: UNIPROT:Q9FTS1; EMBL:AL132972
A:Experimental source: cultivar Columbia; BAC clone T25B15
A:Genetics:
A:Map position: 3
A:Insertions: 122/1; 146/3; 225/2; 277/2; 338/3; 422/2; 535/1; 628/3; 664/3
A>Note: T25B15.80

Query Match	23.9%	Score 800.5	DB 2	Length 737
Beet Local Similarity	31.4%	Pred. No. 4e-48		
Matches 211	Conservative 132	Mismatches 227	Indels 101	Gaps 21
Qy	27	DLKAFTEGAV-----LSFHNLCYRVKLSGFLPCRRKPEKEKELSLNSINGIMKG-	74	
Db	124	DIEAATSSVVFQAPFFPIYLKFDIDITYKTTGG----MTSSSKSLTNGISGAVYGE	179	
Qy	75	LNAILGPTGGKSSLDVLAARKDPGSLGVDLNGABRPANFKNSGVVQDDVMGTL	134	
Db	180	LLAAMGPGSGSKTTLNLNAGRFNQONIGSGSVSYNDKPSYGLHKLTRIGFVQDDVLPFHL	239	
Qy	135	TVRENLQPSALRLATMTNHEKKERINRVIOELGLDQVADSKVGTQGIIRVSGGERKRT	194	
Db	240	TVKSTLLVYTTALLRLPKTLTEQEKQRAASVIOELGLECOOTMIGSGSVRAGSGGERKRV	299	
Qy	195	SIGMELITDPSEILFDEPTTGLDSTANAVLLIKRMSKQRTIIFSIHOPRYSIFKLPD	254	
Db	300	CIGEMINTNPGLLDBERTSLSDTMAKYQMHLCAIAKAGKITVTHIGSSSLFHHFD	359	
Qy	255	SLTLLASGRLMFHGPAOBALGYFESAGVHCEAVNNPADPFLLIDINGDSTAVALNREDFK	314	
Db	360	KLVLASGSLILFGKASEAMSYFSSIGCSPLANNPABFLDLVNGMNDIS-----	411	
Qy	315	ATEIIEBSKQDKPLEKELAEIYVNS-----SFYKETAKEHLQLS-----CGE	356	
Db	412	-----VPSALKEKMKIIRL-ELYVANKCVDETQYLEBAAYKQIAVMEMKMKMAPVLPDEE	466	
Qy	357	KKKKIYAFKE---ISYTTSPFC-HQLRWVSKRS---FKNLGNPOASIAQIIVTVVLGLVI	409	
Db	467	VKLMITPKKEMGMSWEOYICLSLRGIKERRHVFSML-----RTYQVUSTAI---IL	517	
Qy	410	GAIFYGLKNDSTGIQ-NRAGVLFELTNOCFSSV-SAVELFVVEKKLFIEHISIGYVRVS	467	
Db	518	GLLWV---QSDITSGQRPFRSGLFFIAVFWGPFVPTAIFTPQGRAMLKSRBSBNYRLS	575	
Qy	468	SYFLGKLSLDLPMTMLPSIIFCTIYVFMGLKPKADAFPMMTMLMNAVASASMALAI	527	
Db	576	AYFPAARTTSQ-LPDLILPVLFLVWVYFMAGLRLRASFSPFLSVTLFELCIYVMAAGLGIAI	634	
Qy	528	AAGGSVSVAVTLMTTICFVFMMLTFSGLLNVLTTLASLMSLMOYSEIPRYGFTAL---QHN	584	
Db	635	GASIMDLKKAITTLASVYVMTFMLAGYFVK--KVPPFIAMIRFMSFNHYHYKLLVKYOYE	692	
Qy	585	EPFLOQNECPGLNATGNPNCAVATCGEBYLVKQIDISPWGLMKQHVLAACMIVIFLTIA	644	
Db	693	EIM-----ESVNGGE--IESGL-----KEVALVAMITIGRYLVA	724	
Qy	645	YKLKLPFKKYS	655	
Db	725	YFSLRRMKDHS	735	

RESULT 5
C86441
probable ABC transporter [imported] - *Arabidopsis thaliana*

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 03-Jul-2004
C:Accession: C864441
R:Thnologists, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Linn, D.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzman, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-646 <STO>
A:Cross-references: UNIPROT:Q9C6R7; GB:AE005172; NID:g11136734; PIDN:AA63135.1; GSPDB:Q
C:Genetics:
A:Map position: 1
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 23.1%; Score 774; DB 2; Length 646;
Best Local Similarity 33.5%; Pred. NO. 2.4e-46;
Matches 213; Conservative 119; Mismatches 234; Indels 70; Gaps 22;

```

QY 37 LSFHNI CYRKLK--SGFLPCRKVEKEIILNINIMKPG--LNLILPGGGKSSLDLV 93
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 51 ITLKEVYVKKIEOTSOCMSWKSKEKTIILNGITMCPCGEFLMLGPGSGSKTILLSL 110
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 94 AARKDPSGLSDVLI NGAPRPANFKCNSGYVODDVNGTLTVRENLQESAA RLATWT 153
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 111 GGRLSKLT--FSGKVMYNGQPSGCIKRRGRGFAQDDVLVPHLVTWETLFTTLLRLPSSIT 169
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 154 NHEKNERINRVIQELGLDKVADSKVGTQPIFGVSGGEKKRTSIGHELIDPSILFDEPT 213
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 170 RDEKAHVHVDVIAELGNLRCNTSMIGB LFFGDISGEEKRVSIGEMLINPILLDEPT 229
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 214 TGLDSTANAVL LLLKMSKOGRTIIFSIHQPRVYIFPKLFDLTLLAGRLMFGPAQEA 273
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 230 SGLDSTHRIHVIYTTIKRLASGRIYVTTIHQPSRIYHMFDPVLLSBSGPITYGAASA 289
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 274 LGYFESAGYHCEAYNPADFFLDIING--DSTAVALNEBEDKATEIIEPSKODPLEK 331
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 290 VEFSSILGFSTSLVNPA DILLDLANGIPRDTOKETSEQOKTVKETL--VSAYEKNISIK 348
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 332 L-AEI--YVNSFKEKTAELHQLSGGEKKKIIVKELSYTTSFCHQ-----RWSKRS 384
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 349 LKHELNCNASHSYETYPAAAKNL-----KSEOMCTMYOFTVLLQRGVRRR 396
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 385 FKNLLGNPQASIAQIIVTVLGLVIGAIYFGKNDSTGIONRAGVLFELTNOCFSSV--S 443
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 397 FSEF---NKLRIFQVLSVAFLG---GLLW--HHPKSHIQDBTALLFFSVWGEFPLVN 448
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 444 AVELFVVEKKLFHEIYISGYRVSSYPLGKLSIDLLPMTMLPSIIFTCIVYFMGLKPKXA 503
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 449 AVFTFPOEKMKLKERSSGMWRSLSSYFMA RVGD--LPLELALPTAFVFIYMWGGLKDP 507
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 504 DAFVMMFTLMTMYAS---ASSMALATAAGSVSVATLLMTCICVPMWIFSGLVNLT 560
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 508 TTF---ILSLVUVYSLVAOGLGLAGLALMLNITQATTLASVTTLLVFLIAGYVQ--Q 562
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 561 IASWLSMLQYFSIPRYGFTALQHNHFLQONCFPGINATGNPNCAVTC-----GEEY 613
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 563 IPEFIVWLKLYLSYCYCKULL-----GIQYTD--DYEGSGKGVACRGDPE 608
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 614 LVK--QSIDLSPGMLKXHVALLACMIVFLTAIYK 648
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 609 AIKSMGIN---NLMTIDVFWGVMLVGRIMAYML 640
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

096553
hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96553
R:Thelouis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroso, J.S.; Maiti, R.; Marzalli,
Rizzo, M.; Rooney, T.; Rowley, J.R.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; NCID:12016719; PMID:11130712
A:Accession: D96553
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <STO>
A:Cross-references: UNIPROT:Q9C8K2; GB:AE005173; NID:g10092349; PIDN:AA012758.1; GSPDB:C
C:Genetics:
A:Gene: F5D21.6
A:Map position: 1
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-i

[illegible]

RESULT 7

A84509

Probable ABC transporter [imported] - Arabidopsis thaliana

C1:Species: Arabidopsis thaliana (mouse-ear cress)

C2:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #ext_change 09-Jul-2004

C3:Accession: A84509

R1:Lin. X.; Kuhl, S.; Rounleay, S.D.; Shea, T.P.; Bentlo, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, C.M.; Tallon, L.

eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.

Nature 402, 761-768, 1999

A1:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A2:Reference number: A84420; MUID:20083487; PMID:10617197

A3:Accession: A84509

A4:Status: preliminary

A5:Molecule type: DNA

A6:Residues: 1-649 <STO>

A7:Cross-references: UNIPROT:Q9SIR6; GB:AE002093; NID:94558665; PIDN:AAD22683.1; GSPDB:GN

C1:Gene: At2g13610

A:Map position: 2

C2:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-b

Query Match

Best Local Similarity 33.5%; Pred. No. 2, 2e-44; Matches 203; Conservative 116; Mismatches 239; Indels 48; Gaps 18;

61 KEILSININGIMKP-GNALILPTGGKSSLLDLVLAARKDPSGLSGDVLINGAP-REPANK 118

60 KHVLRKQVCRAPWEIILVIGPSGAGKSSLEILARLIQ--TGSVYVKKRPVDRANK 117

119 CNSGVYVODVVMGTLTVENENQFSAALATLMTHEKNERINRIQIOLGDKVADSKV 178

118 KISGVYVQDQTLFPLTVETELLFSAKLL--KLPADELRSRKSLVHLEGLAEVAATAV 175

179 GTQPIRGVSGERKRISIGMELITDPSILFDEPTGLDSTANAVALLKMKSK-QGR 237

176 GDDSVRGISGGERRRISIVIEVHDKVILDEPTSGLDSTSLIITDMLKMAERKGT 235

238 IIFSIHQPRYSIFKLFDSLTLASGLMFGHQAQALGYFESAGYCEAYNNPADFFLDI 297

236 ILTTHQPRFRIVKQNSVLLANGSTLKQSGVDQGVLRNGLHPRLHEIVIEPAIS 295

238 IINGSTAVALNREDFKATEIIEPSKQKPLIEKLAETIVNSFYKETAELHOLSGEK 357

236 I--ESTTKQQRLOESRRRAHVLP--QTLOEKSEDSQGES--KSGFTLLQQLFOQTR 348

358 KKKI--TVPEISYTTSPCH-----QLRWVSKSPKMLGNPQASIQIIVTVVLGVIGAI 412

349 VADVGTMNATEFTDRDFANSRLBETMILTHRPSKNTFTKELPACTVOMLGSGLVGLI 408

413 YFGLKNDSTGIONRAGVLFELTNOCFSSVSAVELFVEEKKLFIHEIYISGYRVSSYFLG 472

409 FNNLADDLKGAERGLFAFILTLSTIETALPIFLOERELIMKETSQSGSRVSSYANA 468

473 KLLSLLPMTMLPSIIFTCIVYFMGLKPKADAFVYMMFTLMVAVSASSMALAIAA--- 529

469 NGLV--YLPELILALIFSTPYVWLVLGNSPFMAFLFSLIIMLIYLTANSVVVCEALVP 527

530 ----GQSVVSVATLMTICFVFMIFPSGLVNLITLASMVLQVRSIRYGFPTALQHE 585

528 NPIVGNVSVISG---VWGSFF-----LPSGFISNHEIPGWIMFHYISLKYIPEEGFLNE 580

586 FLGQNFPCPLNATGNPCNYATCTGEEYLVKQIDLSIPWG---LWKNHVALACMIVFLT 642

581 PFSKMKC---LEGGFGC-----LVTEEDLKE-----ERYGESRMNVVIMCPVLLYMF 629

643 IAYLKL 648

630 ISYVIL 635

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

RESULT 8

T47652

ABC transporter-like protein - Arabidopsis thaliana

N1:Alternate names: protein T26112.10

C1:Species: Arabidopsis thaliana (mouse-ear cress)

C2:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #ext_change 09-Jul-2004

C3:Accession: T47652

R1:Montfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X

submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24471

A2:Accession: T47652

A3:Status: preliminary

A4:Molecule type: DNA

A5:Residues: 1-725 <MON>

A6:Experimental source: cultivar Columbia; BAC clone T26112

C1:Gene: At2g112.10

A:Map position: 3

C2:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-b

Query Match

Best Local Similarity 22.2%; Score 745; DB 2; Length 725; Matches 193; Conservative 130; Mismatches 274; Indels 62; Gaps 13;

36 VLSFHNICRYVRLKSGFLPCRKVEKILSNINGIMKRP-LNAIIGPTGGKSSLLDVLA 94

72 VLNFNLOVDVTLRRRFGFSKQNGVKTLLDDVSGASDDIYLVASGAGSTLIDALA 131

95 ARKDSGSLGVDVLINGAP--REPANKSGVYVODVVMGTLTVENENQFSAALRLATM 152

122 GRVARGSLGSLGVDVLINGAP--REPANKSGVYVODVVMGTLTVENENQFSAALRLATM 191

153 TNEKNERINRIYQELGDKVADSKVGTQPIRGVSGERKRISIGMELITDPSILFDEPT 212

192 SKSKIMBERVEALIDOLGRNANANTVIGDEGHGVSGGERRRISIGIIDIHDPVIFLDP 251

213 TTGLSSTANAVALLKMKMSKQRTIIPSIHQPRYSIFKLFDSLTLASGLMFGHQAQAL 272

252 TSGLDSTNAPVWVQVVKKIAQSGSIVIMSHOPSARIYELLRLITLSKGVFNGSPAS 311

273 ALGYESAGYHCEAYNNPADFFLDI-----NGDSTAVALNREDFKATEIIEPS 322

312 LPGFSDPGRPIPEKENSIFPALDLVRLBEGSNEGTAKLVDNEMKQCKISLIQAPOT 371

333 ---KQDKPLIEKLAETIVNSFYKETAELHOLSGEKKKKITVFEKISYTTSPCHQLRW 379

372 NKLDDRSLSLKEA---INASV---SRGKL--VSGSSRSNPTSMETVSSYANPSLFEFPI 423

380 VSKRSPKMLGNPQASIQIIVTVVLGVIGAIYFGLKNDSTGIONRAGVLFELTNOCF 439

424 LAKRYMKWIRMPBELVGRIATVMTGCLLAVVYKLDHTPRGADERLTLFAFVPTMY 483

440 SSVSAVELFVEEKKLFIHEIYISGYRVSSYFLGKLSLDLPMTMLPSIIFTCIVYFMGL 499

484 CCLDNVPFPIQRIYFLBETTHNAVRTSSYVSHLV--LPDLAPSLVFSAITWTVGL 542

500 KPKADAFVMMFTLMVAVSASSMALAIAAGSVSVATLMTICFVFMIFPSGLVNLIT 559

543 SGLGEGFYVCLLIYASFSGSVVTFISGVVPNIMLCYMAISITLAYCLLSGFEVND 602

560 TIASLMSLVQVRSIRYGFPTALQHEFLQNC-----PGLNAG----- 599

603 RLPFWTWPFHYISILKYPEAVLINEFDDPSKCPVRGQVDFDSTLLGVSQSKVLLT 662

600 ----NNPCNATC--TGEEYLVKQGI--DLSFGLMKHVALACMIVFLTAYVLLP 650

663 LSKSLRTITESTCLARTSGDILAQGITQLSKMD-----CLMITPASGLPFRILP 712

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

N,Alternate names: protein F4P12.210
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C/Accession: T45891
R/Bioecker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quefeler, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A/Reference number: 223016
A/Accession: T45891
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-739 <BLO>
A/Cross-references: UNIPROT:Q91FG8; EMBL:AL132966
A/Experimental source: cultivar Columbia; BAC clone F4P12
C/Genetics:
A/Map position: 3
A/Note: F4P12.210
C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-h

Query Match 22.1%; Score 741.5; DB 2; Length 739;
Best Local Similarity 28.3%; Pred. No. 5.5e-44;
Matches 193; Conservative 123; Mismatches 256; Indels 111; Gaps 17;
36 VLSFHNICVRVKKSGF--LPC--RKPE-----KEILSNINGMKDG-LNALIGPT 82
87 VLSFKDITYSVKIKKKKPPCCGNSPFGNDNEMNTKVLINGSGAREGEMVAIVGAS 146
83 GGGKSLIDVLAARKDPSGLDVLINGADRPANF-KNSGYVVDVDMVMTLTVRENLQ 141
147 GSGKSTLIDALARKISKSRLGDTTLNGEVESLHVISAAYVQDLDLTPMLTVEETM 206
142 FSAALRIATTMTNHEKNERINRVIOELGDKVADSKVGTQFIRGSGEKRTSIGMELI 201
207 FSAEFLRPSLSKKKKARVQALIDQLGRLAAKTIVIGDEGRGVSGERRRVSIIGDII 266
202 TDSIILPDEPTGDSSTANAVLLKRMKSQKRTIIFSIHORYSIFLTFSDITLAS 261
267 HDPIILFDEPTGSDSTAYMVVKVLRQAQSGSIYVMSIHQPSYRILGLDLPLSR 326
262 GRIMFHQPAQALGYFESAGYHCEAVNPPADFLDIINGDSTAVALREEDFKATEIIEP 321
327 GNVYSSPTNHLPOFEBEHPRIENENKPEFALDLR-----ELSDS 369
322 SKODKPLIEKLAETIYVNSFYKETKAEHLHOLSGEKKKKTIVKEI----- 367
370 PEGTKSLIVE-----FKKWRK--QTSQSRNTNVSLLKDAISASISRGKIVSGA 417
368 -----SYTTSCHQLRWVSKSFKNLGNPQASIAQIIVVGLVIGALYFGKND 419
418 TNLRSFOTANPPTWETMLVIGKRSILNSRRQPELFGIRGAVLVMTMLATTIPMKLDS 477
420 STGIQNRAGVLFLLTNQCFSSSAVELFVVEKLFHEIYISGYRVSSYFLGKLSDDL 479
478 PRIGIQLGFFAFAMSTFTYTCALIPVLOEKRIIFRETAIVNAYRRSSIVLAHTIIS-I 536
480 PMTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
537 PALIILISAFAASFFSAVGLAGSEGLPFEFFITLTFAMAGSFVETLGSVSHVIMIGFT 596
540 LMTICFVMMIFSGLVNLTITIASKLSWLOYSIPRIGFTALQNEFLGQNF--PGLNA 597
597 VVAAILAYFLFLFGFIFISDRIDPLVYMWIFHLSLVKPYEGVLQNEEDPTKCVNRIDM 656
598 TGNNPCNYA-----TC--TGEELYVQGI-DLSPMG-LMKNH 630
657 FDSPLQGVPTAVKISLLKMSGVLGINVTAETCVTTGIDILKQGGITBISKNCM--- 713
631 VALACMIVIFLTIA---YKLKLF 650
714 -----ITVAMGFFFRVLV 726

hypothetical protein FSD21.8 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C/Accession: H96552
R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: H96552
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-678 <STO>
A/Cross-references: UNIPROT:Q9C8J8; GB:AE005173; NID:gl0092361; PIDN:AA012770.1; GSPDB:(
C/Genetics:
A/Map position: 1
C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-i

Query Match 22.0%; Score 739; DB 2; Length 678;
Best Local Similarity 30.3%; Pred. No. 7.3e-44;
Matches 186; Conservative 137; Mismatches 248; Indels 42; Gaps 13;
61 KEILSNINGMKDG-LNALIGPTGGKSLIDVLAARKDPSG--LSGDTVLINGADRPAN 116
28 KRLINGNGGCEENRILAIIMPSSGSKSTLIDLALGL--AGVNVSGKLVNKKRRLD 85
117 FKNSGYVVDVDMVMTLTVRENLQPSAALRIATTMTNHEKNERINRVIOELGDKVADS 176
86 FGA-AAVYQEDVULGTLTVRESISYSAHRLSKITREISIVERTITDMGLEBESDR 144
177 KVGTOPIRGVSGEKRTSIGMELITDPSILFLDEPTGDSSTANAVALLKRMKSQGR 236
145 TIGNMHLRGISGEKKRLSTALEVLTKPSLLPDEPTGSDSASAFVVOILNRIASSGK 204
237 TITPSIHQPRYSIFKLPDSITLASGRIMHGAQELGYFESAGYHCEAVNPPADFLD 296
205 TVVSSIHQPSGEVFPALFDLILSGGETYVFGAESATKFGEGAPCPSPRRNPSDHLR 264
297 IINGD-----STAVALNR--EEDFKATEIIEPSKQDKPLIEKLAETIYVNSFYKETKAE 349
265 CVNSDPDNTAALVESRRINDSFSLHQLHETNTIDPL-DIETPAEIRITLVKPKKCSL 323
350 HOLSGEKKKKTIVF-----KEISYTSFCHQLRWVSKSFKNLGNPQASIAQIIVTV 403
324 YAAASRARIGELIASIYIVTERKKSGQTNMVKQLRILTQGSFINMSRDLGYYMMRIAVYI 383
404 VGLGVIGALYFGKANDSTGIQNRAGVLFLLTNQCFSSSAVELFVVEKLFHEIYISGY 463
384 VLSICVGSIFPNVGRNHTNWSSTAACGFAGMTMNSIGFOSFIEEMVYFRERLNGH 443
464 YRVSSYFLGKLSDDLPMTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSASM 523
444 YGVAIVTWSLBS-LPFTILMCLSTSSITTYNWRQSGSHFFYNCLDICALITYBSC 502
524 ALAIAAGQSVSVATLMTIC---FVFMNIFSGLVNLTITIASKLSWLOYSIPRYGFT 579
503 MMMIAS-----VVPNFMGLMGVLAGYIGIVNLVSGGFRFPDLPMTVMWRVPVSYINGAW 557
580 ALQ--HNEFLGQFCPGLNATGNPCNATTCGEEYLVQKQIDLSFGWGLMKHVALACM 636
558 ALQAGYKNEMIGVEY-----DSPLPVKKKGBELIQTIVGINBESSKWLDAVVM 609
637 IVFELTAYLKL 649
610 ILIGRIAPFALL 622

RESULT 11

T47650
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T15C9.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T47650
R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24470
A:Accession: T47650
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <NEW>
A:Cross-references: UNIPROT:Q9M2V5; EMBL:AL132970
A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetic8:
A:Map position: 3
A:Note: T15C9.110
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-b

Query Match 21.7%; Score 726.5; DB 2; Length 708;
Best Local Similarity 30.1%; Pred. No. 5.8e-43;
Matches 198; Conservative 123; Mismatches 282; Indels 55; Gaps 15;
QY 36 VLSFNICRYVLYKGFPCRPVE--KEILSNINGIMKPG-LNALGPTGGKSLDLY 92
DB 62 LLSFNLSYNNVLRFRDFSRRTASVKTLDDIGEADGILAVLGSGAGKSLTDA 121
QY 93 LAARPDGSLGSDVINGAP--RPANFKNSGVVODDVWMTLVRENLOPSAALRLAT 150
DB 122 LAGRAEDELKKTVTTLNGKQVLSRLKTVSAVWODDLFPMILYKETLMASERLR 181
QY 151 TWTNHEKNERINRVIOELGLDVADSKVGTQPIRGVSGGERRKTSIGMELITDPSILFD 210
DB 182 SLPSKKMERVETLLIDQGLRNAADTVIGDEGHRVSGGERRRVSIGIDIHDPILFLD 241
QY 211 EPTTLDSTANAVLLLRMSKQRTTFSHQRPYISFKLPDSITLLASGRLMFHGA 270
DB 242 EPTSLDSTANAVVQVLRKIAQSGVWVMSIHQSPARILIGLLDRILITSHKSVFNGSP 301
QY 271 QEALYFSGAGYHCEAYNNPADFLDI-----NGDSTAVLNREDFKATELIEPSKD 325
DB 302 VSLPSFSSFGRIPEKENITFALDVIRELGSSGTRDIVEFNEKWOQNTARATTS 361
QY 366 KFLIEKLAIEYVNSFFYKETAELHQLSGGEKKKIVFKEISYTTSPFCHQLRWYSKRSF 385
DB 362 RYSLKEAIAASY-----SRGKL--VSGSGGANPISMETVSSYANPPLAEFFILAKRYI 412
QY 386 KULLNPOASTIAQIIVTVVLGLVIGAIYFGLKNDSTGIONRAGVLFELTTNOCFSSVAV 445
DB 413 KMWITPELIGMRIGTVMTGLLATVVRMLNTPRGAQERMGFAFGSMSTFYCCADI 472
QY 446 ELFVVEKKLFIHEYISGYRVASVYFGLKLSDLPMTMLPSIIFTCIYFMLGKPKADA 505
DB 473 PVFIERIYIFLETHNAIRTSYIVISALVS-LPQLMLSTAFAATTFTWIGLSGGLS 531
QY 506 FVVMNFTLMMVAVSASMAIAAQSVSVATLLMTICFV-FMMIFSGLVNLTITIASW 564
DB 532 FFYVCLIIYAAMWSSSIVTPI-SGLIPVMMSYMTIAYLSYCLLGGFYINRRIPLY 590
QY 565 LSMLOYFSPRYGFTALQNEFLGQNF-----PGLNATG 599
DB 591 WIMFYIISLKYRYEAVLINPEFDDPSRCVKGQVFDGTLAEVSHVMKLLDLSGSL 650
QY 600 NNPCNVATC--TGEEVLVKGQI-DLSPMG-LMKNHVALACMIVIFLTIAVLLFLKK 653
DB 651 GTKITESTCLRTGPDLLMQGITTQSKWDCLM---ITLAMGL-FRILFYLSLFLGSK 704

RESULT 12

T08934

hypothetical protein F27G19.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T08934
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16519
A:Accession: T08934
A:Molecule type: DNA
A:Residues: 1-635 <BEV>
A:Cross-references: UNIPROT:Q9S2R9; EMBL:AL078467; GSPDB:GNO0062; ATSP:F27G19.20
A:Experimental source: cultivar Columbia; BAC clone F27G19
C:Genetic8:
A:Gene: ATSP:F27G19.20
A:Map position: 4
A:introns: 38/3; 253/1; 304/1; 414/3
C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 21.5%; Score 721; DB 2; Length 635;
Best Local Similarity 30.0%; Pred. No. 1.2e-42;
Matches 203; Conservative 121; Mismatches 284; Indels 68; Gaps 18;
QY 1 MSSNVEVFIPVSGNTNGFPATASNDLKAFTEGAVLSFHNICRYVLYKSG--FLPCRK 57
DB 1 MDNQESMDVETPIAKTNDRELSPSIFKANNPVLTKFENLVYTVLKDSGCFGKNDK 60
QY 58 PVEKEILSNINGIMKPG-LNALGPTGGKSLDVLAR--KDSGSLGSDVINGAPRP 114
DB 61 TEERTILGLTGIVYKRGILMLGPSGKSLTALAGRVGEGKLTGNSYNKPLS 120
QY 115 ANFKNSGVVODDVWMTLVRENLOPSAALRLATWTNHEKNERINRVIOELGLDKYA 174
DB 121 KAVKTTGFTVODDLYLNYLTETLVFTALRLNSRKQKQIKQAARVMTGLDRK 180
QY 175 DSKVGTQPIRGVSGGERRKTSIGMELITDPSILFDEPTTGLDSTANAVLLLRMSKQ 234
DB 181 DTIGGPFPLRGVSGGERRKVSIGQELINLPSILFDEPTSGDSTTAQRIVIGILWELANG 240
QY 235 GRTTIFSHQRPYISFKLPDSITLLASGRLMFHGAQALGFESAGHCEAYN-NPADF 293
DB 241 GRTVTTTHQF-----SKGNPVYFGGSNAMDYFASGVSPIVERINPSDF 286
QY 294 FLDIINGDSTAVLNREDFKATELIEPSKQKPLIEKLAIEYVNSFFYK-----E 344
DB 287 LLDIANGKPLV-----ISCWPSVGSDEGRPAMKALV---AFKTNLDSVINE 335
QY 345 TKAELHQLSGGEKKKIVFKEISYTTSPFCHQLRWYSKRSFKNLGNPQ--SIAQIIVT 402
DB 336 VKGQDDLGKPKRESRVATNTYGDWPTTWMOQFVLLRGLKQRHDSFGKVAQIF-- 393
QY 403 VVLGLVIGAIYFGLKNDSTGIONRAGVLFELTTNOCFSSV-SAVELFVEKKLFIHEYIS 461
DB 394 -IVSFLCGLLWMOQTK--ISRLODQIGLFFISSFMAFPPLFQOIFTFPOERAMLKERSS 450
QY 462 GYVRYSFELGKLSDLPMTMLPSIIFTCIYFMLGKPKADAFVMMFTLMMVAVSAS 521
DB 451 GMYRLSPYLSKRVGD-LPMELILEPTCFLVITYMMAGLNHLNANFVTLVLVAVLSG 509
QY 522 SMALAIAGOSVSVATLLMTICFVMMIFSGLVNLTITIASWLMLOYFSPRYGFTAL 581
DB 510 GIGLALGLALWMDQKATTLGVSIMLTPLLAGYVYQHVPV-FIMIKVSIQYTYKLL 567
QY 562 QHNEFLGONFCPGLNATGN--PCNVATTCGEEVL-VKQIGDLSFGMLMKNHVALACM 637
DB 568 ----ILGQYTANLELPCDNGKLRCHVDFFGIGIKHIGFNSGL-----VSALATLAML 615
QY 638 VIFLTIAVLLFLKK 653
DB 616 VYRVIAIYALTRIGK 631

RESULT 13

G84791

probable ABC transporter [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C/Accession: G84791

R./lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euse, D.; Nlemaan, W.C.; White, O.; Eise, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A./Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: G84791

A/Status: preliminary

A./Molecule type: DNA

A./Residues: 1-755 <STO>

A./Cross-references: UNIPROT:Q9ZUTO; GB:AE002093; NID:g4056489; PIDN:AAC98055.1; GSPDB:GN

C/Genetics:

A./Map position: 2

C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-b

Query Match

Best Local Similarity 21.5%; Score 720.5; DB 2; Length 755;

Matches 202; Conservative 133; Mismatches 282; Indels 99; Gaps 21;

2 SSSNVEVFPVSGQNTNGPAPATASNDIKAFTEGA-VLSFHNICRYVKLKGFLP--CRK 57

63 SSRALGIASPINBA-ASSFMSWASAPASSISSPFLVTFDLTVSKIKKFNPLACCR 121

58 PVE-----KEISNNGIKKPG-LNALIGPTGGKSLDYLAAKDPGSLGSDVILNG 110

122 SGNDSSVNTKILNGLSGAREGEMAVILASGSGSKTLLDALANRIAKDSLGLSTILNG 181

111 APRPAMP-KCNSGVVODDVVMGTLTVRENIQPSAALRLATTTMNEKNERIRINIOELG 169

182 EVLESSNQKISAYWODDLFPMULTVEETLMSAEERLSKSKKKARQVALDQLG 241

170 LDRVADSKVGTQPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLK 229

242 LRSAAKTVIDEGHGVSGGERRRVSGINDIHDPIITLFDPEFTSGDSTAYVIVKVLQ 301

230 RMSKQRTTIFSHQPIYSIFKLPDSTLLASGRIMHGAQELGFESAGHCEAYNN 289

302 RIAQSGSIVTMSIHQPIYRMGLDQILFISKGTIVSGSPTHLPQFSFEPKIPENEN 361

290 PADPFIIDINGDSTAVANREEDPKATEIIEPSKODPLIE-----KLAETIVN---- 338

362 KTFEALDIL-----RELVS-----TEGTPELVPHKQMAKQAPSYNNKKR 404

339 ----SSFYKETKALHQ--LSGGEKKKKTIVFEKI-SYTTSCHQLRWVSKSFKNLLG 390

405 NTWVSLKEAITASISRGKLVSGATNNNSNLTPSFQTPANPFWIEMIVIGKALINSRR 464

391 NPQASIAQIIVTVVGLVIGAIYFGLKNDSTGIONRGVLPFLTNOCSVSANVELFV 450

465 QPRLGRLGAVWVVTGIIATMTFNLNDSKGAQERGFAPAMSTTFYCAEAIPIVFLQ 524

451 EKLFIHEIYISGYRVSSYFLGKLLSDLPMTLPSTIIFTCIYVFMGLPKADAFVVM 510

525 ERITFMEETAYNAKRSSSYLSOSIIS-IPALVLSAFAATTFWAVGLGANGFFEFY 583

511 FTLMVAVYSASMALAIAQSVSVVATLMTICFVPMIISGLVNLTTIASVLTWLYQ 570

584 FTILASFWAGSSFTVTLISGVIPNVLGFTVVAIIAFLFLFSGFFISRDRIPIYWMLFH 643

571 FSIIPRGFTALQHNPELQON---FCPGLNATGNP-----C 603

644 ISLVKPYEGVNLNEF--QNPTRCFAGVQLFDNSPLGEFPNDVKVNLKMSGVLGTNV 701

604 NVATNC--TGEVYVKGSI-DLSPWG-LMKHVALACIVIFLTA---YKLLF 650

702 TAEFTVTTGIDILKQGGITDLSKNCML-----ITVAGFFFRVLF 742

RESULT 14

102567

probable ATP-binding cassette protein T16B24.1 - Arabidopsis thaliana

N/Alternate names: protein F12L6.1

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C/Accession: T02567; T00545; C84816

R./Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,

submitted to the EMBL Data Library, August 1998

A./Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.

A./Reference number: Z14679

A/Accession: T02567

A/Status: translated from GB/EMBL/DBJ

A./Molecule type: DNA

A./Residues: 1-740 <ROU>

A./Cross-references: UNIPROT:O80946; EMBL:AC004697; NID:g3402671; PIDN:AAC28975.1; PID:g

A/Experimental source: cultivar Columbia

A./Submitted to the EMBL Data Library, July 1998

A./Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.

A./Reference number: Z14168

A/Accession: T00545

A/Status: translated from GB/EMBL/DBJ

A./Molecule type: DNA

A./Residues: 1-362 <ROU>

A./Cross-references: EMBL:AC004218; NID:g3355463; PIDN:AAC27826.1; PID:g3355464

A/Experimental source: cultivar Columbia

R./lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euse, D.; Nlemaan, W.C.; White, O.; Eise, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A./Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: C84816

A/Status: preliminary

A./Molecule type: DNA

A./Residues: 1-740 <STO>

A./Cross-references: GB:AE002093; NID:g3402672; PIDN:AAC28975.1; GSPDB:GN00139

C/Genetics:

A./Map position: 2

C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-i

F/110-310/Domain: ATP-binding cassette homology <ABC>

Query Match

Best Local Similarity 21.4%; Score 717.5; DB 1; Length 740;

Matches 193; Conservative 120; Mismatches 294; Indels 83; Gaps 16;

26 NDLKAFTEGAVLSFHNICRYVKLK-----SGFLPCR-----KPYEKELSNING 69

56 NDGYMRTVPVPSFNDLTVVSVRPKLDPRRLTPRRRTEDPEIAQTPARPKTKILLNINS 115

70 IMKPG-LNALIGPTGGKSLDYLAAKDPGSLGSDVILNAGAPRAN-PKNSGVVOD 127

116 ETRDGEIMVILGASGSGKSTLLDALANRIAKGSLKGTIVKNGTLLQSRMLKVASAYMOD 175

128 DVVMGTLTVRENIQPSAALRLATTTMNEKNERINVOELGADKVAASKVGTQIRGV 187

176 DLTFPMLTVEETIMFAERFLPSRLPSKSKKLVQALIDQLGRNAKTIIGDEGHRGIS 235

188 GGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLKMSKQRTIIFSHQPRY 247

236 GGERRVSIGIDILHPIILFLDEPISGSDSTAFVWVKVLRKIAQSGSIVMSIHQPSH 295

248 SIFKLPDSTLLASGRIMHGAQELGFESAGHCEAYNPADPFLDI-----NG 300

296 RVLGDLDRILFISRGHTVVSGPSASLPREFTEGSPIPENENTETALDIRELEAGAG 355

301 DSTAVANREEDPKATEIIEPSKODPLIEKLAETIVNSFYKETVAKELHQ--LSGGEK 357

356 TRGLIEPNK---KQWEMKQSNRPPLTPP-SSPYNLTUKAIAIASISRGKLVSGGES 410


```

Qy 358 -----KKKIVFEKISYTTSPFCHOLRWYSKRSFKNLGNPQASIAQIIVTVVLGIVGA 411
      |          :          :          :          :          :          :
Db 411 VAHGATNTTTLTAFAFANPMMEIKITLSKSMNSRRQPELFGIRASVITGTILAT 470
      |          :          :          :          :          :          :
Qy 412 IYFGLKNDSTGIQNRAGVLPFLTNOCFSSVASVELFVVEKKLFIEHYISGYRVSYYL 471
      |          :          :          :          :          :          :
Db 471 VEMRLDNSPKGVQOERIGFPAFAMSTWTFYCADALPVFQBERYIFMEETAYNNVRRSSYYL 530
      |          :          :          :          :          :          :
Qy 472 GKLLSDLLPMTLPSIIFTCIYFVMLGKPKADAFVMMFTLMVAVSASMAALAAQ 531
      |          :          :          :          :          :          :
Db 531 SHAIYS-FPSLLIFLSVAFAATTWAVGLDGLTGLLFCILILASFSGSSFTFTLSGV 589
      |          :          :          :          :          :          :
Qy 532 SVASVATLMTICFVPMIFSGILVNLTTIASVLSWLOYSIPRYGFTALQHNHFLQNF 591
      |          :          :          :          :          :          :
Db 590 PSMVLCYTIIVAILAVFLFSGFFINRNIPDWMIFHMSLVKRYEAVLQNEFSDARK 649
      |          :          :          :          :          :          :
Qy 592 C--PGINATGNP-----CNYATC--TGEYLVKQG-IDLSP 623
      |          :          :          :          :          :          :
Db 650 CFVRGVQIPDNTPLGELPEVMKLLGTVSKSLGVITISSTCLTGTSDILRQGVVQLSK 709
      |          :          :          :          :          :          :
Qy 624 WGLMKHVALACMIVIELTIA---YKLIF 650
      |          :          :          :          :          :          :
Db 710 KN-----CLFITVAFGFFPRILF 727

```

RESULT 15

```

B96742
Probable ABC transporter F17M19.11 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 09-Jul-2004
C:Accession: E96742
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Anen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talon,
Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; WUID:21016719; PMID:11130712
A:Accession: E96742
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-609 <STO>
A:Cross-references: UNIPROT:Q9C8W6; GB:AE005173; NID:G6978921; PIDN:AAF34313.1; GSPDB:GN
C:Genetics:
A:Gene: F17M19.11
A:Map position: 1
C:Superfamily: Fruit fly white protein; ATP-binding cassette homology

```

```

Query Match      21.3%; Score 712.5; DB 2; Length 609;
Best Local Similarity 31.0%; Pred. No. 4.5e-42;
Matches 190; Conservative 125; Mismatches 256; Indels 41; Gaps 16;

```

```

Qy 56 RKPVEKEILSNNGIMKPG-LNALICPTGGKSSLDVLAARKDPSGLSGDVLINGAPRP 114
      |          :          :          :          :          :          :
Db 23 RSTERTILISGYTGMISPEFMAVLGPSSGKSTLINAAVAGRLHGSNLTGKILINDGKIT 82
      |          :          :          :          :          :          :
Qy 115 ANFKKSGVYVDDVVMGTLTVRENQFSAALRLATTMTNHEKNERINRVIOELGDKVA 174
      |          :          :          :          :          :          :
Db 83 KQTLKRTGTVADDDLIYPLHIVRETLVFAALLRLPRSLTRDYKLAESVISELGLTKE 142
      |          :          :          :          :          :          :
Qy 175 DSKVGTQPIRGVSGERKRTSIGMELITPSILFLDEPTTGLDSTANAVALLLKMSK- 233
      |          :          :          :          :          :          :
Db 143 NTVVGNTPFRIGSGGRKRVSIABELLINPSLVLDEPTSGDATAALRLVOTLAGLHG 202
      |          :          :          :          :          :          :
Qy 234 QGRITIIFSIHQRYSTIFKLPDSLTLASGRLMFHGPAQDALGYFESAGYHCEAYNNPADF 293
      |          :          :          :          :          :          :
Db 203 KQKTVTSTIHQSSRVFQMFDTVLLSEKCLFVGKGRDAMAYFESVGPSPAPFPMNPAF 262
      |          :          :          :          :          :          :
Qy 294 FLIDING--DSTAVALNREDFKATEIIPSPKODKPLIEKLAIEI---YVNSFYKETKA 347

```

```

Db 263 ILDDLNGVCOQTDGVTREKERNVRQTLVTRAYDTLLAPQVTCIEVSHFPQDNARFVXT--- 319
      |          :          :          :          :          :          :
Qy 348 ELHQLSGEKKKKKIVFEKISYTTSPFCHOL-RWYSKRSFKNLGNPQASIAQIIVTVVLG 406
      |          :          :          :          :          :          :
Db 320 ---RVNGGGITTCIA-----TWFSQLCILHLRLKRRHSEF-----DLRIFOYVAAS 365
      |          :          :          :          :          :          :
Qy 407 LVIGAIYFGLKNDSTGIQNRAGVLPFLTN--OCFSSVASVELFVVEKKLFIEHYISGYR 465
      |          :          :          :          :          :          :
Db 366 ILICGLMWM--HSDYRDVHDLGLLFFISIFMGVLSFNNAVFTFPQERAFTEBRASGWT 423
      |          :          :          :          :          :          :
Qy 466 VSSYFGLKLLSDLLPMTLPSIIFTCIYFVMLGKPKADAFVMMFTLMVAVSASMA 525
      |          :          :          :          :          :          :
Db 424 LSSYFMAHVLGSLSMELVLPASFLT-FTYMMVYLRPGIYVFLTLVLLVYLASQGLG 482
      |          :          :          :          :          :          :
Qy 526 AIAAGOSVSVATLMTICFVPMIFSGILVNLTTIASVLSWLOYSIPRYGFTALQ 582
      |          :          :          :          :          :          :
Db 483 ALGAAIMPAKKAISTIVYTMALFVLTGGYVA--KVPSSGMVMKVVSTTFYCYRLVAIQ 540
      |          :          :          :          :          :          :
Qy 583 HNEFLGQNFPC--GLNATGNPCNYATCTGEYLVKQIGIDLSPMGLMKHVALACMIVIF 640
      |          :          :          :          :          :          :
Db 541 YGS--GEELRMVLGCDSSKKGKQASAAITSAGCFVEEVI--GDVGMTSVGVLFMLPFQY 596
      |          :          :          :          :          :          :
Qy 641 LTIAYLKLFLK 652
      |          :          :          :          :          :          :
Db 597 RVLAYLALRRIK 608

```

```

Search completed: June 6, 2005, 09:44:30
Job time : 47 secs

```

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2005, 09:19:13 ; Search time 175 Seconds
(without alignments)
1916.639 Million cell updates/sec

Title: US-09-961-086-1
Perfect score: 3352
Sequence: 1 MSSSVEVFIPVSGQNTNGF.....MIVFLTIATYKLPLKKYS 655

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trcembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3346	99.8	655	2	Q96TA8
2	3342	99.7	655	2	Q81X16
3	3339	99.6	655	1	ABG2_HUMAN
4	3338	99.6	655	2	Q96LDB
5	2849.5	85.0	656	2	Q8MIB3
6	2762	82.4	657	2	Q7TM55
7	2761	82.4	541	2	Q86V64
8	2757	82.2	657	2	Q9R004
9	2754	82.2	657	2	Q80W57
10	2752	82.1	657	2	Q80ST1
11	2742	81.8	657	2	Q80XFT
12	1787.5	53.3	650	2	Q8BK15
13	1786.5	53.3	650	1	ABG3_MOUSE
14	1703.5	50.8	646	2	Q68HW7
15	864	25.8	1039	2	Q68IH1
16	863	25.7	645	2	Q68G61
17	847.5	25.3	801	2	Q8T691
18	835.5	24.9	1049	1	ADP1_YEAST
19	826.5	24.7	1011	2	Q756R4
20	821.5	24.5	1078	2	Q7SHS0
21	819	24.4	1055	2	Q6FKY0
22	812	24.2	687	1	WHIT_DROME
23	809	24.1	751	2	Q93Y54
24	808	24.1	687	2	Q9NH94
25	806	24.0	687	2	Q94960
26	802.5	23.9	1022	2	Q6CS24
27	800.5	23.9	737	2	Q9FT51
28	793	23.7	679	2	Q9BH97
29	790	23.6	692	2	P91892
30	783.5	23.4	648	2	Q9C6M5
31	777.5	23.2	679	2	Q81S30

32	776.5	23.2	567	2	Q9FGL7	Q9FGL7 arabidopsis
33	774	23.1	646	2	Q9C6R7	Q9C6R7 arabidopsis
34	772	23.0	670	2	Q77423	Q77423 bacteroides
35	772	23.0	679	1	WHIT_CERCA	Q17320 ceratitidis
36	767.5	22.9	687	2	Q9C8R2	Q9C8R2 arabidopsis
37	766.5	22.9	695	2	Q81M05	Q81M05 oryza sativ
38	763.5	22.8	798	2	Q8T689	Q8T689 dictyostel
39	759	22.6	703	2	Q8RXN0	Q8RXN0 arabidopsis
40	755	22.5	672	2	Q91182	Q91182 arabidopsis
41	754.5	22.5	677	1	WHIT_LUCCU	Q05360 lucilia cup
42	751	22.4	669	2	Q8WRP2	Q8WRP2 tribolium c
43	751	22.4	695	1	WHIT_ANOGA	Q27256 anopheles g
44	749.5	22.4	691	2	Q8RW19	Q8RW19 arabidopsis
45	746	22.3	649	2	Q981T6	Q981T6 arabidopsis

ALIGNMENTS

RESULT 1
Q96TA8 PRELIMINARY; PRT; 655 AA.
AC Q96TA8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE ATP-binding cassette superfamily G (white) member 2 (ATP-binding cassette, sub-family G, member 2).
GN Name=ABG2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21201983; PubMed=11306452;
RA Komatani H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,
RA Arakawa H., Nishimura S.;
RT "Identification of breast cancer resistant protein/mitoxantone
RT resistance/placenta-specific, ATP-binding cassette transporter as a
RT transporter of NB-506 and J-107088, topoisomerase I inhibitors with an
RT indolocarbazole structure.";
RL Cancer Res. 61:2827-2832 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rabe S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalski U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ABC transporter family.

DR EMBL; AB051855; BAB46933.1; -.
DR EMBL; BC021281; AAH21281.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042826; F:ATPase activity, coupled to transmembrane m. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 655 AA; 72313 MW; ABAF6B96034C5A8 CRC64;

Query Match 99.8%; Score 3346; DB 2; Length 655;
Best Local Similarity 99.8%; Pred. No. 1.2e-208;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEFLPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRPVE 60
DB 1 MSSSNVEFLPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRPVE 60
QY 61 KEILSININGIMKGLNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120
DB 61 KEILSININGIMKGLNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120
QY 121 SGYVODDVMGTLTYRENLOFSALRLATTTMNEKNERINRIVIOELGLDKVADSKVGT 180
DB 121 SGYVODDVMGTLTYRENLOFSALRLATTTMNEKNERINRIVIOELGLDKVADSKVGT 180
QY 181 QPFRGVSGBERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240
DB 181 QPFRGVSGBERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240
QY 241 SHOPRYSIFKLPDSITLLASGRLMFHGAOELAGFESAGYHCEAYNNPADFFLIING 300
DB 241 SHOPRYSIFKLPDSITLLASGRLMFHGAOELAGFESAGYHCEAYNNPADFFLIING 300
QY 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEYVNSFPYKETAELHQLSGGEKKK 360
DB 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEYVNSFPYKETAELHQLSGGEKKK 360
QY 361 ITYFKKISTTSCHOLRWKRSKSPKULLGNPOASIAQIIVTVVLGLVIGAIYFGKND 420
DB 361 ITYFKKISTTSCHOLRWKRSKSPKULLGNPOASIAQIIVTVVLGLVIGAIYFGKND 420
QY 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEYISGYRRVSSYFLGLSLDLLP 480
DB 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEYISGYRRVSSYFLGLSLDLLP 480
QY 481 MTMLPSIIFTCIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540
DB 481 MTMLPSIIFTCIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540
QY 541 MTICPVFMNIFSGLLVNLTTIASWLSWLYFSIPRGFTALQHNELGONFCGLNATGN 600
DB 541 MTICPVFMNIFSGLLVNLTTIASWLSWLYFSIPRGFTALQHNELGONFCGLNATGN 600
QY 601 NPCNVAATCGEEVLYVQGLIDSPWGLMKONVALACMIVIFLTAIVKTLFLKKYS 655
DB 601 NPCNVAATCGEEVLYVQGLIDSPWGLMKONVALACMIVIFLTAIVKTLFLKKYS 655

RESULT 2
081X16 PRELIMINARY; PRT; 655 AA.
AC 081X16;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ATP-binding cassette protein ABCG2.
GN Name=ABCG2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA Yoshikawa M., Yabuchi H., Ikegami Y., Ishikawa T.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -i- SIMILARITY: Belongs to the ABC transporter family.
EMBL; AF463519; AA014617.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 655 AA; 72314 MW; ABAF6B591D4C5A8 CRC64;

Query Match 99.7%; Score 3342; DB 2; Length 655;
Best Local Similarity 99.7%; Pred. No. 2.2e-208;
Matches 653; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEFLPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRPVE 60
DB 1 MSSSNVEFLPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRPVE 60
QY 61 KEILSININGIMKGLNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120
DB 61 KEILSININGIMKGLNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120
QY 121 SGYVODDVMGTLTYRENLOFSALRLATTTMNEKNERINRIVIOELGLDKVADSKVGT 180
DB 121 SGYVODDVMGTLTYRENLOFSALRLATTTMNEKNERINRIVIOELGLDKVADSKVGT 180
QY 181 QPFRGVSGBERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240
DB 181 QPFRGVSGBERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240
QY 241 SHOPRYSIFKLPDSITLLASGRLMFHGAOELAGFESAGYHCEAYNNPADFFLIING 300
DB 241 SHOPRYSIFKLPDSITLLASGRLMFHGAOELAGFESAGYHCEAYNNPADFFLIING 300
QY 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEYVNSFPYKETAELHQLSGGEKKK 360
DB 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEYVNSFPYKETAELHQLSGGEKKK 360
QY 361 ITYFKKISTTSCHOLRWKRSKSPKULLGNPOASIAQIIVTVVLGLVIGAIYFGKND 420
DB 361 ITYFKKISTTSCHOLRWKRSKSPKULLGNPOASIAQIIVTVVLGLVIGAIYFGKND 420
QY 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEYISGYRRVSSYFLGLSLDLLP 480
DB 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEYISGYRRVSSYFLGLSLDLLP 480
QY 481 MTMLPSIIFTCIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540
DB 481 MTMLPSIIFTCIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540
QY 541 MTICPVFMNIFSGLLVNLTTIASWLSWLYFSIPRGFTALQHNELGONFCGLNATGN 600
DB 541 MTICPVFMNIFSGLLVNLTTIASWLSWLYFSIPRGFTALQHNELGONFCGLNATGN 600

0y	601	NPCCNATCTGGEYLVKQSIDLSPGCMKGNHVALACMIYIPLTIALTKLFLPKKS	655
Db	601	NPCCNATCTGGEYLVKQSIDLSPGCMKGNHVALACMIYIPLTIALTKLFLPKKS	655
RESULT 3			
ABG2_HUMAN	STANDARD.	PRT.	655 AA.
ID	ABG2_HUMAN	Q9UN001; Q95374; Q9B773; Q9UN001;	
AC	Q9UN001	Q95374 (Rel. 40, Created)	
DT	16-OCT-2001	(Rel. 40, Created)	
DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DT	05-JUL-2004	(Rel. 44, Last annotation update)	
DE	ATP-binding cassette, sub-family G, member 2 (placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).		
GN	Name=ABCG2; Synonyms=ABCP, BCRP, BCRL1;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OC	NCBI_TaxID=9606;		
RN	[1]	SEQUENCE FROM N.A.	
RP	TISSUE=Placenta;		
RC	MDLINE=99065313; PubMed=9850061;		
RX	Allikmetz R., Schirral L.M., Hutchinson A., Romano-Spica V., Dean M.,		
RT	"A human placenta-specific ATP-binding cassette gene (ABCP) on		
RT	chromosome 4q22 that is involved in multidrug resistance.";		
RL	Cancer Res. 53:5337-5339 (1998).		
RP	[2]	SEQUENCE FROM N.A.	
RC	TISSUE=Breast cancer;		
RX	MDLINE=99080071; PubMed=9861027; DOI=10.1073/pnae.95.26.15665;		
RA	Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,		
RT	Roos D.D.;		
RT	"A multidrug resistance transporter from human MCF-7 breast cancer		
RT	cells.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670 (1998).		
RN	[3]	ERRATUM.	
RP	Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,		
RA	Roos D.D.;		
RA	Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569 (1999).		
RP	[4]	SEQUENCE FROM N.A.	
RC	Kage K., Tsuchihara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.		
RA	Sugimoto Y.;		
RT	"Breast cancer resistance protein constitutes a 140-KDa complex as a		
RT	homodimer.";		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
RN	[5]	SEQUENCE OF 198-655 FROM N.A.	
RP	TISSUE=Placenta;		
RX	PubMed=14702039; DOI=10.1038/ng1285;		
RA	Ora T., Suzuki Y., Nishikawa T., Otsubi T., Sugiyama T., Irie R.,		
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,		
RA	Sehara M., Ohashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,		
RA	Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,		
RA	Shiraori A., Mutsaers K., Yasuda T., Iwatsunagi T., Wagatsuma M.,		
RA	Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,		
RA	Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.		
RA	Omura Y., Abe K., Kamihara K., Katsuba N., Sato K., Tanikawa M.,		
RA	Yamazaki M., Nishimura K., Ishibashi T., Yamashita K., Murakawa K.,		
RA	Fujimori K., Tanai H., Kimata M., Watanabe M., Hirao S., Chiba Y.,		
RA	Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotta T.,		
RA	Kusano Y., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,		
RA	Nomura Y., Togiyasu S., Komai F., Hara R., Takeuchi K., Arita M.,		
RA	Imobe N., Mueselino K., Yuuki H., Oshima A., Sasaki N., Aotsuma S.,		
RA	Mochikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,		
RA	Yoshioka S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,		
RA	Nishigaki S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Makebe H.,		
RA	Hibagaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,		
RA	Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,		
RA	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,		
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,		

RA	Kawabata.., Hikiji T., Kobateke N., Inagaki H., Ikema Y., Okamoto S.,
RA	Okltani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA	Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA	Togashi T., Oyama M., Hata H., Matsunabe M., Komatsu T.,
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT	"Complete sequencing and characterization of 21,243 full-length human
RT	cDNAs.";
RL	Nat. Genet. 36:40-45 (2004).
RN	[6]
RN	REVIEW.
RP	MEDLINE=21474438; PubMed=11590207;
RA	Schmitz G., Langmann T., Heimerl S.;
RT	"Role of ABCG1 and other ABCG family members in lipid metabolism.";
RL	J. Lipid Res. 42:1513-1520(2001).
RN	[7]
RP	VARIANTs LEU-431 AND LEU-489.
RA	Itoda M., Saito Y., Shirao K., Minami H., Ohtsu A., Yoshida T.,
RA	Saijo N., Suzuki H., Sugiyama Y., Ozawa S., Sawada J.-I.;
RT	"Eight novel single nucleotide polymorphisms in ABCG2/BCRP in Japanese
RT	cancer patients administered irinotecan.";
RL	Drug Metab. Pharmacokinet. 18:212-217 (2003).
CC	-I- FUNCTION: Xenobiotic transporter that appears to play a major role
CC	in the multidrug resistance phenotype of a specific MCF-7 breast
CC	cancer cell line. When overexpressed, the transfected cells become
CC	resistant to mitoxantrone, daunorubicin and doxorubicin, display
CC	diminished intracellular accumulation of daunorubicin, and
CC	manifest an ATP-dependent increase in the efflux of rhodamine 123.
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC	-I- SIMILARITY: Belongs to the ABC transporter family. ABCG (white
CC	subfamily).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements/ab-bib.ch).
CC	-- or send an email to license@ab-bib.ch).
CC	-----
DR	EMBL; AF103796; AAD09188.1; -
DR	EMBL; AF098951; AAC97367.1; -
DR	EMBL; AB056867; BAB39212.1; -
DR	EMBL; AK002040; BAA92050.1; -
DR	Genew; HGNC:74; ABCG2.
DR	MI: 603756; -
DR	GO; GO:0016021; C:Integral to membrane; TAS.
DR	GO; GO:0005524; F:ATP binding; TAS.
DR	GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . ; TAS.
DR	GO; GO:0005215; F:transporter activity; TAS.
DR	GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.
DR	GO; GO:0042493; P:response to drug; TAS.
DR	GO; GO:0006810; P:transport; TAS.
DR	InterPro; IPR003593; AAA_ATPase.
DR	InterPro; IPR003439; ABC_transporter.
DR	Pfam; PF00005; ABC_tran; 1.
DR	ProDom; PD000006; ABC_transporter; 1.
DR	SMART; SMO0182; AAA; 1.
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW	ATP-binding; Polymorphism; Transmembrane; Transport.
FT	DOMAIN 1 395
FT	TRANSSEM 396 416
FT	DOMAIN 417 428
FT	TRANSSEM 429 449
FT	DOMAIN 450 477
FT	TRANSSEM 478 498
FT	DOMAIN 499 506
FT	TRANSSEM 507 527
FT	DOMAIN 528 535
FT	TRANSSEM 536 556
FT	DOMAIN 557 630
FT	Extracellular (Potential).

```
FT TRANSMEM 631 651 Potential.
FT DOMAIN 652 655 Cytoplasmic (Potential) .
FT NP_BIND 80 87 ATP (Potential) .
FT CARBOHYD 418 N-linked (GlcNAc . .) (Potential) .
FT CARBOHYD 557 557 N-linked (GlcNAc . .) (Potential) .
FT CARBOHYD 557 557 N-linked (GlcNAc . .) (Potential) .
FT VARIANT 431 431 F -> L.
FT VARIANT 431 431 /Ftrid=VAR_018349.
FT VARIANT 489 489 F -> L.
FT VARIANT 489 489 /Ftrid=VAR_018350.
FT CONFLICT 24 24 V -> A (in Ref. 2 and 4) .
FT CONFLICT 166 166 E -> Q (in Ref. 2 and 4) .
FT CONFLICT 208 208 F -> S (in Ref. 1) .
FT CONFLICT 315 316 Missing (in Ref. 5) .
FT CONFLICT 482 482 R -> T (in Ref. 2) .
SQ SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CCE0 CRC64;

Query Match 99.6%; Score 3339; DB 1; Length 655;
Best Local Similarity 99.5%; Pred. No. 3,4e-208;
Matches 653; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSSNVEVFLPVSGQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPE 60
DB 1 MSSSNVEVFLPVSGQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPE 60
QY 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
DB 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
QY 121 SGVYVDDVVMGTLTVRENLOFSALRLATTTMNHKNERINRVIOELGLDYADSKVGT 180
DB 121 SGVYVDDVVMGTLTVRENLOFSALRLATTTMNHKNERINRVIOELGLDYADSKVGT 180
QY 181 QFIRGVSGGRRKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKMSKQGRITIF 240
DB 181 QFIRGVSGGRRKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKMSKQGRITIF 240
QY 241 SHQPRYSIFKLPDSLTLLASGRLMFHGPAQOELGFESAGYHCEAYNNPADPFLDING 300
DB 241 SHQPRYSIFKLPDSLTLLASGRLMFHGPAQOELGFESAGYHCEAYNNPADPFLDING 300
QY 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
DB 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
QY 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPQASIAQIIVTVVLGIVGAIFYGLKND 420
DB 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPQASIAQIIVTVVLGIVGAIFYGLKND 420
QY 421 TGIQNRAGVLFPLTNNQCFSSVSAVELFVVEKELFIEHYSIGYRVSSYFLGKLSDLLP 480
DB 421 TGIQNRAGVLFPLTNNQCFSSVSAVELFVVEKELFIEHYSIGYRVSSYFLGKLSDLLP 480
QY 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540
DB 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540
QY 541 MTICFPMFMTFSGLLVNLITIASWLSWLOFYFSIPRYGFTALQNHFEIGQNFCELANATGN 600
DB 541 MTICFPMFMTFSGLLVNLITIASWLSWLOFYFSIPRYGFTALQNHFEIGQNFCELANATGN 600
QY 601 NPCNVATCTGEELYVKGIDLSPMGLKMNVALACMIVIFLITAYLLKLLKKS 655
DB 601 NPCNVATCTGEELYVKGIDLSPMGLKMNVALACMIVIFLITAYLLKLLKKS 655

RESULT 4
Q96LD6 PRELIMINARY; PRT; 655 AA.
AC Q96LD6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
```

```
DE ABC transporter ABCG2.
GN Name=ABCG2;
OS Homo sapiens (human) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21424790; Pubmed=11533706;
RA Zhou S., Schuetz J.D., Bunting K.D., Colapietro A.M., Sampath J.,
RA Morris J.J., Lagutina I., Grosveld G.C., Osawa M., Nakauchi H.,
RA Sorrentino B.P.;
RT "The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of
RT stem cells and is a molecular determinant of the side-population
RT phenotype."
RT Nat. Med. 7:1028-1034(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AY017168; AAG52982.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: GO:0000166; F:nucleotide binding; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003459; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Ppatine_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00893; ABC_TRANSPORTER 2; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR KW ATP-binding.
SQ SEQUENCE 655 AA; 72287 MW; B3B5DC02095C4A8 CRC64;

Query Match 99.6%; Score 3338; DB 2; Length 655;
Best Local Similarity 99.7%; Pred. No. 4e-208;
Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSSNVEVFLPVSGQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPE 60
DB 1 MSSSNVEVFLPVSGQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPE 60
QY 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
DB 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
QY 121 SGVYVDDVVMGTLTVRENLOFSALRLATTTMNHKNERINRVIOELGLDYADSKVGT 180
DB 121 SGVYVDDVVMGTLTVRENLOFSALRLATTTMNHKNERINRVIOELGLDYADSKVGT 180
QY 181 QFIRGVSGGRRKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKMSKQGRITIF 240
DB 181 QFIRGVSGGRRKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKMSKQGRITIF 240
QY 241 SHQPRYSIFKLPDSLTLLASGRLMFHGPAQOELGFESAGYHCEAYNNPADPFLDING 300
DB 241 SHQPRYSIFKLPDSLTLLASGRLMFHGPAQOELGFESAGYHCEAYNNPADPFLDING 300
QY 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
DB 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
QY 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPQASIAQIIVTVVLGIVGAIFYGLKND 420
DB 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPQASIAQIIVTVVLGIVGAIFYGLKND 420
QY 421 TGIQNRAGVLFPLTNNQCFSSVSAVELFVVEKELFIEHYSIGYRVSSYFLGKLSDLLP 480
DB 421 TGIQNRAGVLFPLTNNQCFSSVSAVELFVVEKELFIEHYSIGYRVSSYFLGKLSDLLP 480
```

Oy		481	MTMLSIIFTCTVYFNLGKPKRADAFVMWFLLTMVAASSSMALIAAGOSVVSATTL	540
Dd		481	NRMLSIIFTCIVTFMGLAKADAFVMMFTLMMVAISSSALLAIAQSVSAVTLL	540
Oy		541	MTICVFPMWIPESGLLVNLTIIASWLSWLOFYSPIRYGPTALQHNEFLGONFCPLNATGN	600
Dd		541	MTICGVFMWIPESGLLVNLTIIASWLSWLTQYFSIPRGGFTALQHNEFLGONFCPLNATGN	600
Oy		601	NPCNVATCTGEERYLYKQIGIDLSPWGMKNHVALACHTIVFLTAYIKLFLKKYS	655
Dd		601	NPCNVATCTGEERYLYKQIGIDLSPWGMKNHVALACHTIVFLTAYIKLFLKKYS	655
RESULT 5				
QMIB3				
ID	QMIB3	PRELIMINARY;	PRT;	656 AA.
AC	QMIB3;			
DT	01-OCT-2002 (TREMBLrel. 22, Created)			
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)			
DT	01-MAR-2004 (TREMBlrel. 26, Last annotation update)			
DE	Brain multidrug resistance protein.			
GN	Name=BMDP;			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
CC	NCBI_TaxId=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22050127; PubMed=12054514; DOI=10.1016/S0006-291X(02)00376-5;			
RA	"Eisenblatter T., Galla H.J."			
RT	"A new multidrug resistance protein at the blood-brain barrier."			
RL	Biochem. Biophys. Res. Commun. 293:1273-1278(2002)."			
CC	-I- SIMILARITY: Belongs to the ABC transporter family.			
DR	EMBL; AJ242927; CADI2785.1; -.			
DR	PIR; JC7860; JC7860.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:000524; F:ATP binding; IEA.			
DR	GO:0042628; F:ATPase activity; coupled to transmembrane m. . .; IEA.			
DR	GO:0000168; F:nucleotide binding; IEA.			
DR	GO:0006810; F:transport; IEA.			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	InterPro; IPR006162; Pntantne_S.			
DR	Pfam; PF00005; ABC_tran.1.			
DR	ProDom; PD000006; ABC_transporter; 1.			
DR	SMART; SM00382; AAA_1.			
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.			
DR	PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.			
KV	ATP-binding.			
SO	SEQUENCE	656 AA;	72391 MW;	118ADSBS3D9D67F CRG64;
Query Match 85.0%; Score 2849.5; DB 2; Length 656;				
Best Local Similarity 84.3%; Pred. No. 2e-176;				
Matches 553; Conservative 44; Mismatches 58; Indels 1; Gaps 1				
Oy		1	MSSSNVEVPFIPISOGNTNGFPATASNLDCAFPNEGAVLSFNHCYRYKLKSGFLPCRKPVE	60
Dd		1	MSSSNVQYVIPSKSKNTNGLPSSSHNELTTSAGCAVLASHDDCYRYKVASGGLPCRKTIVE	60
Oy		61	KEILSNINGIMKPGNALIIGPTGGGKSLLDVLAARKDSGLSDVLINGAPRAPFKCN	120
Dd		61	KEILTININGIMKPGNALIIGPTGGGKSLLDVLAARKDPHGJSGDVLINGAPRAPFKCN	120
Oy		121	SGYVVQDDVWGTLVRENLFQSALARLATTTMTNHENKRINRVIOELGIDKYADSKVGT	180
Dd		121	SGYVVQDDVWGTLVRENLQFSAALRIPTTMTNHENKRINIMVIOELGIDKYADSKVGT	180
Oy		181	QPIRGVSGGERKRTSIGMELTDPSILFDPEPTTGDSSTANAVLLLKRMSKGRTIIF	240
Dd		181	QPIRGVSGGERKRTSIAMELITDPSILFDPEPTTGDSSTANAVLLLKRMSKGRTIIF	240
Oy		241	SIHQPRYSIFKFLDSLTLIASGRLEFHGAQBALGYFESAGYHCEAYNNPADFLDIING	300

Db	241	SIHQPRYSIIFKLFDSLTTLASGRLLPHFGAREALGFIASIGVNCBEPYNNADFFLDVING	3000
Qy	301	DSTAVALNR-BEDFPATETIEIBSPKDDPLIEKLAIEIVYVSSFPKETAELHQSGEKKK	3580
Db	301	DSSAVLSRADPDEBAGPEEPPEPEKDTPLIDLTAAPYTNSSFPKOTKVELDQSGGRKKK	3660
Qy	360	KITVEKEISYTTSPCHQLRWMSKSPFKNLGNPDASIAOIIVTVLGLVIGAIYFGKND	4190
Db	361	KSSVYKEVYTTTSPFCHQLRWMSKSPFKNLGNPDASIAOIIVTVLGLVIGAIYFGKND	4200
Qy	420	STGIONRAGVLPFLTNOCFFSSVSAVELFVWRKKLPFIHYISGYVYVSSYFGKLSDDL	4790
Db	421	PSGIONRAGVLPFLTNOCFFSSVSAVELLVWRKKLPFIHYISGYVYVSSYFGKLSDDL	4800
Qy	480	PMTMLPSLIIFTCIVYFMGLKPKKADAFVYWMFTLLMWVAASASSMALAIAGOSVAVTL	5390
Db	481	PMRMLPSLIIFTCIVYFMGLKPKKADAFVYWMFTLLMWVAASASSMALAIAGOSVAVTL	5400
Qy	540	LMTICFVFMWIFPSGLLVNLTITIASWLSWLYOFSIPRYGFTALQHNFFLQCNFCPGINATG	5990
Db	541	LMTISFVFMWIFPSGLLVNLTITIASWLSWLYOFSIPRYGFTALQHNFFLQCNFCPGINATG	6000
Qy	600	NNPCNYACCTGGEYLVKQIGIDLSPMGIMGNHVALACMIYIFLTIALTKLLFLKKYS	6550
Db	601	NNTCFPAICTGAEYLENQGISLSAWGLMGNHVALACMWYIFLTIALTKLLFLKKYS	6560
RESULT 6			
Q7TMS5	Q7TMS5	PRELIMINARY;	PRT; 657 AA.
AC	Q7TMS5;		
DT	01-OCT-2003 (Tremblrel. 25, Created)		
DT	01-OCT-2003 (Tremblrel. 25, Last sequence update)		
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)		
DE	ATP-binding cassette, sub-family G, member 2.		
GN	Name=Abcg2;		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;		
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;		
RA	Stratberg R.L., Feingold E.A., Grouse L.H., Derge J.G.		
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,		
RA	Alteich L.F., Zeeberg B., Buecaw K.H., Schefer C.F., Hsieh N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marushina K., Farmer A.M., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casavut T.L., Schetz T.E.,		
RA	Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J.,		
RA	Boeck S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,		
RA	Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Kozminski A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;		
RA	Stratberg R.L.		
RA	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: Belongs to the ABC transporter family.		
DR	EMBL, BC053730; AAH53730.1; --		
GO	GO:0016021; C:integral to membrane; TAS.		

DR InterPro: IPR003593; AAA ATPase.
DR InterPro: IPR003439; ABC transporter.
DR InterPro: IPR006162; Pnabne_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD00006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KM ATP-binding.
SQ SEQUENCE 657 AA; 72977 MW; DCD70C5D9FA2B5AF CRC64;

Query Match 82.4%; Score 2762; DB 2; Length 657;
Best Local Similarity 81.5%; Pred. No. 9.4e-171;
Matches 536; Conservative 52; Mismatches 66; Indels 4; Gaps 3;

QY 1 MSSNNVEFIPVQSGNTNGPPATASNDLKAFTEGAVLSFNICRYKVLKSGFPCRPVE 60
DB 1 MSSNNHVLVPMQSRNNNGLPRTNSRAVRLAGDVLSPFHITRYRVKVGKGL-VRKIVE 59
QY 61 KEILSNINGIMKPGNALIGPTGSGKSLDVLAAKDPGSLGSDVLINGAPPANPKCN 120
DB 60 KEILSDINGIMKGLNAILGPTGSGKSLDVLAAKDPGSLGSDVLINGAPPANPKCC 119
QY 121 SGYVODDVVMGTLTVRENLQFSALRLATMTNHEKNERINRYIOELGLDKVADSKVGT 180
DB 120 SGYVODDVVMGTLTVRENLQFSALRLPTTMKHEKNERINTLIKELGLEKVAADSKVGT 179
QY 181 QPFRGVSGGRRKRTSGMEILTPPSILFDEPTTGDDSTANAVLLLLKMSKQGRITIF 240
DB 180 QPFRGVSGGRRKRTSGMEILTPPSILFDEPTTGDDSTANAVLLLLKMSKQGRITIF 239
QY 241 SIHQPRYSIFKLFDLSLTLLASGRLMFHPAOEALGYESAGYCEAVNPPADFFLDING 300
DB 240 SIHQPRYSIFKLFDLSLTLLASGRLVFRGPQKALEYASGYCEPNNPADFFLDVING 239
QY 301 DSTAVLANREE-DFKATELIEPSKODKPLIEKLAETIVNSSFYKETVAELHQLSGEKKK 359
DB 300 DSAVVMNREQNEANKTEPSEKPEVIEINLSEFYINSALYGETVAELDQLPGAQEEK 359
QY 360 KIVPEKISITTSFCHQLRVWSKSPKLLGNPOASIAQIIVTVVLGLVIGALYFGKND 419
DB 360 GTSAFKEPVYVTSFCHQLRWIARRSPFNLLGNPOASVAQDIIVTLLIGALYFDLKYD 419
QY 420 STGIQRAGVLFPLTNQCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGLKLSDL 479
DB 420 AAGMKRAGVLFPLTNQCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGRKMSDL 479
QY 480 PMTMLPSIIFTCIYVFMGLKPKRDAEFVMMFTLMVAVYASSSMALAAAGSVSVATL 539
DB 480 PMFLPSVIFTCVLYFMGLKKTVDAFIMFTLIMVAYASSSMALAAAGSVSVATL 539
QY 540 LMTICVFPMVIFSGLLVNLTTIASVLSWLOYSIPRGFTALQHNELGONFCGLNATG 599
DB 540 LMTIAVFPVMMIFSGLLVNLTTIGPWSLWLOYSIPRGFTALQNEFLGDFGCFGVNTD 599
QY 600 NNPC-NYATCGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTIAVLLFLKXY 655
DB 600 NSTCNSYALCTGNEYLINGIELSPWGLMKNHVALACMIIIFLTIAVLLFLKXY 657

RESULT 7
Q86V64 PRELIMINARY; PRT; 541 AA.
AC Q86V64;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ATP-binding cassette sub-family G member 2 (Fragment).
GN Name=ABCG2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22959505; Pubmed=12358161;
RA Zhang W., Mojilovic-Petrovic J., Andrade M.F., Zhang H., Ball M.,
RT Stanimirovic D.B.;
RT "The expression and functional characterization of ABCG2 in brain
endothelial cells and vessels";
RL FASEB J. 17:2085-2087(2003).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AY288307; AAP31310.1; -!
DR GO: GO:001620; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Pnabne_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD00006; ABC_transporter; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KM ATP-binding.
FT NON TER 1 1
SQ SEQUENCE 541 AA; 60450 MW; 050B2742C6A3F66C CRC64;

Query Match 82.4%; Score 2761; DB 2; Length 541;
Best Local Similarity 99.8%; Pred. No. 8.6e-171;
Matches 540; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 115 ANFKNSGVVODDVVMGTLTVRENLQFSALRLATMTNHEKNERINRYIOELGLDKVA 174
DB 1 ANFKNSGVVODDVVMGTLTVRENLQFSALRLATMTNHEKNERINRYIOELGLDKVA 60
QY 175 DSKVQPFIRGVSGGRRKRTSGMEILTPPSILFDEPTTGDDSTANAVLLLLKMSKQ 234
DB 61 DSKVQPFIRGVSGGRRKRTSGMEILTPPSILFDEPTTGDDSTANAVLLLLKMSKQ 120
QY 235 GRTIIFSIHQPRYSIFKLFDLSLTLLASGRLMFHPAOEALGYESAGYCEAVNPPADFF 294
DB 121 GRTIIFSIHQPRYSIFKLFDLSLTLLASGRLMFHPAOEALGYESAGYCEAVNPPADFF 180
QY 295 LDIINDSTAVLANREDEFKATEIIEPSKODKPLIEKLAETIVNSSFYKETVAELHQLSG 354
DB 181 LDIINDSTAVLANREDEFKATEIIEPSKODKPLIEKLAETIVNSSFYKETVAELHQLSG 240
QY 355 GEKKKKTIVPEKISITTSFCHQLRVWSKSPKLLGNPOASIAQIIVTVVLGLVIGALYF 414
DB 241 GEKKKKTIVPEKISITTSFCHQLRVWSKSPKLLGNPOASIAQIIVTVVLGLVIGALYF 300
QY 415 GLKNDSTGIQNRAGVLFPLTNQCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGLK 474
DB 301 GLKNDSTGIQNRAGVLFPLTNQCFSSVSAVELFVVEKKLFHEYISGYRVSSYFLGLK 360
QY 475 LSDLLPMTMLPSIIFTCIYVFMGLKPKRDAEFVMMFTLMVAVYASSSMALAAAGOSVY 534
DB 361 LSDLLPMTMLPSIIFTCIYVFMGLKPKRDAEFVMMFTLMVAVYASSSMALAAAGOSVY 420
QY 535 SVATLMTICVFPMVIFSGLLVNLTTIASVLSWLOYSIPRGFTALQHNELGONFCPG 594
DB 421 SVATLMTICVFPMVIFSGLLVNLTTIASVLSWLOYSIPRGFTALQHNELGONFCPG 480
QY 595 LNAIGNNPNYATCGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTIAVLLFLKXY 654
DB 481 LNAIGNNPNYATCGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTIAVLLFLKXY 540

RESULT 8
Q9R004 PRELIMINARY; PRT; 657 AA.
AC Q9R004;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, last sequence update)
 DT 01-MAY-2004 (TRENBLrel. 26, last annotation update)
 DE Breast cancer resistance protein 1.
 OS Name=Abcg2; Synonyms=Bcrpl;
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB; TISSUE=Liver;
 RX MEDLINE=99413474; PubMed=10485464;
 RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
 RT "The mouse Bcrpl/Mxr/Abcg2 gene: amplification and overexpression in
 cell lines selected for resistance to topotecan, mitoxantrone, or
 doxorubicin."
 RT Cancer Res. 59:4237-4241(1999).
 RL -1- SIMILARITY: Belongs to the ABC transporter family.
 CC EMBL, AF140218; AAD54216.1; -.
 DR MGD; MGI:1347061; Abcg2.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR006162; Pnante_5.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 DR ATP-binding.
 SQ SEQUENCE 657 AA; 73021 MW; 207B70BC272CC0D5 CRC64;

Query Match 82.2%; Score 2757; DB 2; Length 657;
 Best Local Similarity, 81.5%; Pred. No. 2e-170; Indels 4; Gaps 3;
 Matches 536; Conservative 51; Mismatches 67;
 QY 1 MSSNVEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCKRPVE 60
 DB 1 MSSNDHVLVPMSSQNNKGLPMSBRAVTTAEGDVLSTHHTYKVKVSGFL-VRKTYE 59
 QY 61 KEILSNINGIMKPGINALILGPTGGKSSLDVLAARKDPSGVLINGARPPANPKCN 120
 DB 60 KEILSDINGIMKPGINALILGPTGGKSSLDVLAARKDPSGVLINGARPPANPKCC 119
 QY 121 SGYVVDVVMGTLVRENLOPSAALRLATMTNHEKNERIRVVOELGLDVAASKYGT 180
 DB 120 SGYVVDVVMGTLVRENLOPSAALRLPTTMKNEKNERINTTIIKEIGLEKVAADSKYGT 179
 QY 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLLKRMSKQGRITIF 240
 DB 180 QIRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLLKRMSKQGRITIF 239
 QY 241 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 300
 DB 240 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 299
 QY 301 DSTAVALLNR-DFKATEIIEPSKODKPLIEKLAETIYNNSSFYKTKAEHLQSGEKKK 359
 DB 300 DSSAVMLNRREODNANKTEEPSKGEKPYIEMLSERTYNSAIYGETKELDLDPAQOEKK 359
 QY 360 KITVKEISYTTSPFCHOLRWVSKSPKNLGNPOASIQIITVTVLGLVIGVIGLKN 419
 DB 360 GTSARKEPYVTSFCHOLRWVSKSPKNLGNPOASIQIITVTVLGLVIGVIGLKN 419
 QY 420 STGIONRAGVLFELTTNQCFSVSAVELFVVEKKLFIHEIYISGYRVSSYFLGKLLSDL 479
 DB 420 AAGMGNRAGVLFELTTNQCFSVSAVELFVVEKKLFIHEIYISGYRVSSYFLGKLLSDL 479
 QY 480 PWTMLPSIITFCIVFPMGLKPKADAFPMFTLMMVAVSASSMALALIAAGSVSVATL 539
 DB 480 PWRFLPSVITFCILFPMGLKKTVDAAFIMFTLLIMVAVTASSMALALATGOSVSVATL 539
 QY 540 LMTICFVFMIMISGLLVNLTIASMLSWLQYFSIDRYGFTALQHNBFIGNPCGLNATG 599

DB 540 LMTICFVFMIMISGLLVNLTIASMLSWLQYFSIDRYGFTALQYNBFLGQFCPGFNVD 599
 QY 600 NNPC--NATCGCEGYLTKOGIDLSPKMLKRNHVALACIVFELTAYLKLFLKYS 655
 DB 600 NSTCVNSYAICGNEYLLINOGLIESPWLGMKNHVALACIITIFLTAYLKLFLKYS 657
 RESULT 9
 Q80W57 PRELIMINARY; PRT; 657 AA.
 AC Q80W57;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, last sequence update)
 DT 01-MAY-2004 (TRENBLrel. 26, last annotation update)
 DE ABC transporter ABCG2.
 GN Name=Abcg2;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mistar;
 RA Hori S., Ohtsuki S., Tachikawa M., Kimura N., Kondo T., Watanabe M.,
 RA Nakashima E., Terasaki T.;
 RT "Functional expression of rat ABCG2 on the luminal side of brain
 capillaries and its enhancement by astrocyte-derived soluble
 factor(s)."
 RT J. Neurochem. 90:526-536(2004).
 RL -1- SIMILARITY: Belongs to the ABC transporter family.
 CC EMBL; AB105817; BAC76396.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0042626; F:ATPase activity; IEA.
 DR GO; GO:0006166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR006162; Pnante_5.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 DR ATP-binding.
 SQ SEQUENCE 657 AA; 72960 MW; C975C61A08489027 CRC64;
 Query Match 82.2%; Score 2754; DB 2; Length 657;
 Best Local Similarity, 81.0%; Pred. No. 3.1e-170; Indels 4; Gaps 3;
 Matches 533; Conservative 52; Mismatches 69;
 QY 1 MSSNVEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCKRPVE 60
 DB 1 MSSNDHVLVPMSSQNNKGLPMSBRGARTLAEGDVLSTHHTYKVKVSGFL-VRKTYE 59
 QY 61 KEILSNINGIMKPGINALILGPTGGKSSLDVLAARKDPSGVLINGARPPANPKCN 120
 DB 60 KEILSDINGIMKPGINALILGPTGGKSSLDVLAARKDPSGVLINGARPPANPKCS 119
 QY 121 SGYVVDVVMGTLVRENLOPSAALRLATMTNHEKNERIRVVOELGLDVAASKYGT 180
 DB 120 SGYVVDVVMGTLVRENLOPSAALRLPKAMKTHEKNERINTTIIKEIGLEKVAADSKYGT 179
 QY 181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLLKRMSKQGRITIF 240
 DB 180 QIRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLLKRMSKQGRITIF 239
 QY 241 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 300
 DB 240 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 299
 QY 301 DSTAVALLNR-BEDFKATEIIEPSKODKPLIEKLAETIYNNSSFYKTKAEHLQSGEKKK 359

```
Db 300 DSSAVMLNRGEQHANKTEEPSREKPIIENLAEPFINSITTYGETAELDQLPVAKKK 359
Qy 360 KITVFEISYTTSPCHQLRWVSKRSFKNLGNPOASIAQIIVTVGLVIGAIYFGKND 419
Db 360 GSSAFREPVVYTSFCHQLRWIARSPFNLLGNPOASIAQIIVTVGLVIGAIYFGKND 419
Qy 420 STGIQNRAGVLFLLTNOCFSSVAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Db 420 PTGMQNRAGVFFLLTNOCFSSVAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Qy 480 PMTMLPSIIIFTCIVYFMGLGKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
Db 480 PMRFLPSVIYTCILYFMGLGKRTVEAFIMFTLIMVAYASSMALAIAAGQSVSVATL 539
Qy 540 LMTICFVFMVIFSGLLVNLTTIASWLSWLOFSPRIGFTALQHNFLQGNFCPLNATG 599
Db 540 LMTISFVFMVIFSGLLVNLTTIGWLSWLOFSPRIGFTALQHNFLQGNFCPLNATG 599
Qy 600 NNPC--NYATCTGEEVLVKQIGIDLSPWGLMKRNHVLAACMIVITLTAIYKLLFLKYS 655
Db 600 NSTCVNSYITCTGNDYILINQIGIDLSPWGLMRNVHVALACMIIIFLTIAYLKLFLKYS 657
```

RESULT 10

Q80ST1 PRELIMINARY; PRT; 657 AA.

```
AC Q80ST1;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2004 (TReMBLrel. 27, Last annotation update)
DE ATP-binding cassette protein G2 transcript variant B (ATP-binding
DE cassette protein G2 transcript variant C) (ATP-binding cassette
DE protein G2 transcript variant A).
GN Name=Abcg2;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Yabuchi H., Ishikawa T.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AY089996; AAM09106.1; -.
DR EMBL: AY089997; AAM09107.1; -.
DR EMBL: AY089998; AAM09108.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: GO:0000166; F:nucleotide binding; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Pntane_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA_1.
DR PROSITE: PSS0893; ABC_TRANSPORTER_2; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR ATP-binding.
KW ATP-binding.
SQ SEQUENCE 657 AA; 72960 MW; E194871E1C1AC201 CRC64;
```

Query Match 82.1%; Score 2752; DB 2; Length 657;

Best Local Similarity 80.9%; Pred. No. 4.2e-170;

Matches 532; Conservative 53; Mismatches 69; Indels 4; Gaps 3;

```
Qy 1 MSSSNVEVPIPVGQNTNGFPATASNDLKAFTEGAVLSFNNICRYVLLKSGFLPCRKPV 60
Db 1 MSSSNHVLVPMQGRNKGFLPGKSSRGARLAEQDVLSFHITRYRVKSGFL-VKRTAE 59
Qy 61 KEILSINGIMKGLNALIIGPTGGKSSLLDVLAARKDPGSLGVDVINGAPRPANFKCN 120
```

```
Db 60 KEILSINGIMKGLNALIIGPTGGKSSLLDVLAARKDPGSLGVDVINGAPRPANFKCS 119
Qy 121 SGVYVDDVVMGLVYRENLFPSAALRLATMTNHEKNERINRVIQELGDKYADSVGT 180
Db 120 SGVYVDDVVMGLVYRENLFPSAALRLPRAMTKHEKNEINRVIQELGDKYADSVGT 179
Qy 181 QFTRVSGGERKRTSIGMELITDPSILFDEPTGDSSTANAVALLLKMSKQRTIIF 240
Db 180 QFTRVSGGERKRTSIGMELITDPSILFDEPTGDSSTANAVALLLKMSKQRTIIF 239
Qy 241 SHQPRYSIFKLPDSITLLASGLMFRGPAQELGFPESSGHCYCEAYNNPADFFLDING 300
Db 240 SHQPRYSIFKLPDSITLLASGLMFRGPAQKLEFFASAGYCEAYNNPADFFLDING 299
Qy 301 DSTAVVLMNR-EEDFKATTEIIEPSKODKPLEKAEIYVNSPFKETKAEHLQSGEKKK 359
Db 300 DSSAVMLNRGEQHANKTEEPSREKPIIENLAEPFINSITTYGETKAELDQLPVAKKK 359
Qy 360 KITVFEISYTTSPCHQLRWVSKRSFKNLGNPOASIAQIIVTVGLVIGAIYFGKND 419
Db 360 GSSAFREPVVYTSFCHQLRWIARSPFNLLGNPOASIAQIIVTVGLVIGAIYFGKND 419
Qy 420 STGIQNRAGVLFLLTNOCFSSVAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Db 420 PTGMQNRAGVFFLLTNOCFSSVAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Qy 480 PMTMLPSIIIFTCIVYFMGLGKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
Db 480 PMRFLPSVIYTCILYFMGLGKRTVEAFIMFTLIMVAYASSMALAIAAGQSVSVATL 539
Qy 540 LMTICFVFMVIFSGLLVNLTTIASWLSWLOFSPRIGFTALQHNFLQGNFCPLNATG 599
Db 540 LMTISFVFMVIFSGLLVNLRTIGWLSWLOFSPRIGFTALQHNFLQGNFCPLNATG 599
Qy 600 NNPC--NYATCTGEEVLVKQIGIDLSPWGLMKRNHVLAACMIVITLTAIYKLLFLKYS 655
Db 600 NSTCVNSYITCTGNDYILINQIGIDLSPWGLMRNVHVALACMIIIFLTIAYLKLFLKYS 657
```

RESULT 11

Q80XF3 PRELIMINARY; PRT; 657 AA.

```
AC Q80XF3;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE ATP-binding cassette transporter ABCG2.
GN Name=Abcg2;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimizu K., Satake M., Okaya A., Kitahara J., Kitahara N.,
RA Takemura M., Sakagami M., Terada N., Tsujimura T.;
RT "Hepatic Oval Cells Have the Side Population Phenotype Defined by
RT Expression of ATP-binding Cassette Transporter ABCG2/BCRP1.";
RL Am. J. Pathol. 0:0-0(2003).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AB094089; BAC75666.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: GO:0000166; F:nucleotide binding; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Pntane_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA_1.
```

DR	PROSITE: P550893; ABC_TRANSPORTER_2; 1.
DR	PROSITE: P580012; PHOSPHOTRANSFERASE; UNKNOWN_1.
KM	ATP-binding.
SO	SEQUENCE 657 AA; 72961 MW; 458980CC3903D5CE CRC64;
Query Match	81.8%; Score 2742; DB 2; Length 657;
Best Local Similarity	80.9%; Pred. No. 1,9e-169;
Matches 532; Conservative 51; Mismatches 71; Indels 4; Gaps 3	
QY	1 MSSSNEVEFI PVSOGNTNGFPATASNDLAFTEGAVLSFHNICYRKLSGFLPCRKYE 60
DB	1 MSSSDHVLVPMSSQNNKGLPGMSSSGARTLAEGDVLSTHHITRYRKVSGFL-VAKTAE 59
QY	61 KEILSNINCIMKPGNLALIGPTGGKSSLLDVLAAKDPDGLSGDVLNGAPRPANFKN 120
DB	60 KEILSDINGIMKPGNALIGPTGGKSSLLDVLAAKDPDGLSGDVLNGAPQANFKS 119
QY	121 SGVVVDDVDMVLTLYRENLQFSALRLATMTNHNKNRIRNVIOELGDVADSKGT 180
DB	120 SGVVVDDVDMVLTLYRENLQFSALRLAKMTNHNKNRIRNTIYKELGDVADSKGT 179
QY	181 QFIRVSGGGRKRTSIGMELIDPSTLFDDEPTTGDSSTANAVLLLRMSKOGRTIIF 240
DB	180 QFTRISIGGERKRTSIGMELIDPSTLFDDEPTTGDSSTANAVLLLRMSKOGRTIIF 239
QY	241 SIHQPRYSIFKLPDSLTLLASGRLMFHGAQELGYFESAGYHCEAYNNPADFLDIING 300
DB	240 SIHQPRYSIFKLPDSLTLLASGRLMFHGAQALAEYFAGAYHCEYNNPADFLDIVING 299
QY	301 DSTAAVALNR-EEDFPAKTELIEPSKODKPLIEKLAETLYVNSSPFYKTEARLHDSGEEK 359
DB	300 DSSAVALNRGEODHEANKTEEPSKKEKPIENLAETYNSTIYGETKALDOLPVAQKK 359
QY	360 KITVEKEISYTTSPFCHQLRWVSKRSPKNLGNPQASIAQIIVTVVGLVIGALYFGKND 419
DB	360 GSPFKEPEPYVTSFCHQLRWIARRSPKNLGNPQASVADIVTVIGLITGALYFELKND 419
QY	420 STGIQNRAGVLEFLTTNQCFSVSAAVELFVEYKULFIHEYISGYRVSSYFLGKLSDLL 479
DB	420 PTGMQNRAGVLEFLTTNQCFTSVSAVELFVEYKULFIHEYISGYRVSSYFPGKLVSDLL 479
QY	480 PWTMLPSLIIFTCTIVYFMGLKRAKADAFPMVMTLMWVAYSASSMALAIAAGOSVSVATL 539
DB	480 PWRPLPSVIYITCILYFMGLKRLVEAFPIMRFTLLINVAYTASSMALAIAAGOSVSVATL 539
QY	540 LMTIFCVPMFIPSGLLVNLTTTASWLSWQYSPISPYGTALOHNEFLGQNCPCGNATG 599
DB	540 LMTISFVPMFLPSGLLVNLRITIGPMLSWQYSPISPYGTALOHNEFLGQNCPCGNATVM 599
QY	600 NNPC-NYATCTGEEYLVKQIGIDLSPWGLMKHVALACIVIFLTIAYLKLFLTKYS 655
DB	600 NSTCNSATITCTGNYLYLNGIGIDLSPWGLMRHVALACIIIFLTITAYLKLFLTKYS 657
RESULT 12	
Q8BK15	PRELIMINARY; PRT: 650 AA.
AC	Q8BK15;
DT	01-MAR-2003 (TREMBLrel. 23, Created)
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE	Mus musculus 12 days embryo eyball cDNA, RIKEN full-length enriched library, clone: D3304AK03 product: ATP-binding cassette, sub-family G (WHITE), member 3, full insert sequence.
DE	(WHITE), member 3, full insert sequence.
CN	Name=Abcg3;
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_Taxid=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Eyeball;
KX	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

RA Carninci P., Hayashizaki Y.;
RT High-efficiency full-length cDNA cloning.";
Mech. Enzymol. 303:19-44(1999).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA The FANTOM Consortium,
the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
Genome Res. 10:1617-1630(2000).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
Suni N., Ishi Y., Nakamura S., Hazama M., Nishie T., Harada A.,
Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuda S., Kawai J.,
Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
Genome Res. 10:1757-1771(2000).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
Hoti F., Imocani K., Ishi Y., Itoh M., Kagawa I., Kasukawa T.,
Kato H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
Nishi K., Nomura K., Numazaki R., Ono M., Ohnato N., Okazaki Y.,
Saito R., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
Sato K., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ database.
CC -1. SMITAIRTY: Belongs to the ABC transporter family.
DR EMBL: AK051880; BAC34799.1; --
DR MGI: 1351624; Abcg3.
DR GO: 0016021; C: integral to membrane; TAS.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding.
SQ SEQUENCE 650 AA; 73612 MW; B9959BDBDFD945E1 CRC64;
Query Match 53.3%; Score 1787.5; DB 2; Length 650;
Best Local Similarity 55.3%; Pred. No. 1,6e-107;
Matches 363; Conservative 104; Mismatches 174; Indels 15; Gaps 5;
1 MSSANVEVFIPVSGQNTNGEPATASNDLKAFTEGAVLSFNNICRYVKLSGGLPCKRPVE 60

```

Db 1 MASNNDPTVISMIERHLCDEPTNTSDLTTLTEBAVLSFHNIISYQETVSGFPLRKKAYV 60
Qy 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRPANPKCN 120
Db 61 IERLSNISGIMKPGALNAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRPANPKCT 120
Qy 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Qy 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Qy 181 QPFRVSGGGRKRSIGMELITDPSILFLDEPTGDSSTANAVALLLKXMSKQRTIIF 240
Db 174 QNIRPKSKELRKRKTSIAMELVTEHPILFLDDPTGDLRTTDDVILRRKSKGRTIIF 233
Qy 241 SIHOPRYSIFKLPDSLTLLASGRLMFHPAQOALGFYESAGYHCEAVNNPADPFLDIING 300
Db 234 SINOPQYSIFKLPDSLTLLASGRLMFHPAQOALGFYESAGYHCEAVNNPADPFLDIING 293
Qy 301 DSTAVALNRE---EDFKATEIIEPSKODKPLIEKLAIEIYVNSFYKETKALHQLSGGK 357
Db 294 GFSNIIDTEBDGHEDDKYEELEFEROY---VTGKLAMNYAQSPLYSETRAIIIDQLGEOK 350
Qy 358 KKKITVKEISYTTSEFCHOLRWVSKSRFNLLGNPQASIAQIIVTVVLGVLGAIYGLK 417
Db 351 LERSASV-ETTCVTPFCHOLRWVSKSRFNLLGNPQASIAQIIVTVVLGVLGAIYGLK 409
Qy 418 NDSTGIONRAGVLFELTTCNOCSSVAVELFVVEKLFHEIYSGYRVSSYFLGKLSD 477
Db 410 NDCEVQMRAGLYLTLTFCITSVSAGELFVIDRVPFLHEHTSGYRVSSYFLGKLAE 469
Qy 478 LLEPTMLPSIIFTCIYFEMGLKPKADAFVMMFTLMMVAVSASMAALIAAGSVVSA 537
Db 470 LIPRLIPSTVFSILITVIAGVMSKCFETMCTIIVLAVSASPLISIGAGENAVAVP 529
Qy 538 TLLMTICFVMMIFSGLLVNLTTIASWLSMLQYFSIRYGFALQHNEFLGNFCGLNA 597
Db 530 TLLVTVFVPMLEFSGSLVSGSFLPKLSIYQFSIRYGFALQHNEFLGNFCGLNA 589
Qy 598 TGNNPC-NYATCTGEEYLVKQGLIDSPMGMLKNNVALACMIVFLTAVALKLEFLK 652
Db 590 EYVSRCHNYICTGEFLMTOGIDLSMGFWENHLAVCMIIITLTVQLOQV 645

RESULT 13
ABG3 MOUSE STANDARD; PRT; 650 AA.
AC 099P8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE ATP-binding cassette, sub-family G, member 3.
GN Name=Abcg3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen; PubMed=11178751;
RX MEDLINE=21030753;
RA Mckley L., Jain P., Miyake K., Schriml L.M., Rao K., Fojo T.,
RT "An ATP-binding cassette gene (ABCG3) closely related to the multidrug
transporter ABCG2 (MXR/ABCP) has an unusual ATP-binding domain.";
RL Mamm. Genome 12:86-88(2001).
CC -1- SUBUNIT: May dimerize with another subunit to form a functional
transporter.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: Highest levels of expression in thymus and
spleen. Detected in lung and small intestine.
CC -1- SIMILARITY: Belongs to the ABC transporter family. ABCG (white)
subfamily.
CC -1- CAUTION: Seems to have a defective ATP-binding region.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF324242; AKL14241.1; -.
DR MGI; MGI:1351624; Abcg3.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Transmembrane, Transport.
FT DOMAIN 1 387 Cytoplasmic (Potential).
FT TRANSMEM 388 408 1 (Potential).
FT DOMAIN 409 420 Extracellular (Potential).
FT TRANSMEM 421 441 2 (Potential).
FT DOMAIN 442 469 Cytoplasmic (Potential).
FT TRANSMEM 470 490 3 (Potential).
FT DOMAIN 491 498 Extracellular (Potential).
FT TRANSMEM 499 519 4 (Potential).
FT DOMAIN 520 527 Cytoplasmic (Potential).
FT TRANSMEM 528 548 5 (Potential).
FT DOMAIN 549 623 Extracellular (Potential).
FT TRANSMEM 624 644 6 (Potential).
FT DOMAIN 645 648 Cytoplasmic (Potential).
SQ SEQUENCE 650 AA; 73623 MW; 86A5AB4DD26971C CRC64;

Query Match 53.3%; Score 1786.5; DB 1; Length 650;
Best Local Similarity 55.3%; Pred. No. 1.9e-107;
Matches 363; Conservative 104; Mismatches 174; Indels 15; Gaps 5;

Qy 1 MSSNNEVFIPIVSGNTNGPAPASNDLKAFTGAVLSFHNIISYQETVSGFPLRKKAYV 60
Db 1 MASNNDPTVISMIERHLCDEPTNTSDLTTLTEBAVLSFHNIISYQETVSGFPLRKKAYV 60
Qy 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRPANPKCN 120
Db 61 IERLSNISGIMKPGALNAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRPANPKCT 120
Qy 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Qy 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRVIQELGDKVADSKVGT 180
Qy 181 QPFRVSGGGRKRSIGMELITDPSILFLDEPTGDSSTANAVALLLKXMSKQRTIIF 240
Db 174 QNIRPKSKELRKRKTSIAMELVTEHPILFLDDPTGDLRTTDDVILRRKSKGRTIIF 233
Qy 241 SIHOPRYSIFKLPDSLTLLASGRLMFHPAQOALGFYESAGYHCEAVNNPADPFLDIING 300
Db 234 SINOPQYSIFKLPDSLTLLASGRLMFHPAQOALGFYESAGYHCEAVNNPADPFLDIING 293
Qy 301 DSTAVALNRE---EDFKATEIIEPSKODKPLIEKLAIEIYVNSFYKETKALHQLSGGK 357
Db 294 GFSNIIDTEBDGHEDDKYEELEFEROY---VTGKLAMNYAQSPLYSETRAIIIDQLGEOK 350
Qy 358 KKKITVKEISYTTSEFCHOLRWVSKSRFNLLGNPQASIAQIIVTVVLGVLGAIYGLK 417
Db 351 LERSASV-ETTCVTPFCHOLRWVSKSRFNLLGNPQASIAQIIVTVVLGVLGAIYGLK 409
Qy 418 NDSTGIONRAGVLFELTTCNOCSSVAVELFVVEKLFHEIYSGYRVSSYFLGKLSD 477
Db 410 NDCEVQMRAGLYLTLTFCITSVSAGELFVIDRVPFLHEHTSGYRVSSYFLGKLAE 469
Qy 478 LLEPTMLPSIIFTCIYFEMGLKPKADAFVMMFTLMMVAVSASMAALIAAGSVVSA 537
Db 470 LIPRLIPSTVFSILITVIAGVMSKCFETMCTIIVLAVSASPLISIGAGENAVAVP 529
Qy 538 TLLMTICFVMMIFSGLLVNLTTIASWLSMLQYFSIRYGFALQHNEFLGNFCGLNA 597

```

Db 530 TLVLTIVYFVFMFLFSGLSLYPSPFLPKSLWIOYFSIPHYGFRALLNNEFLGONFCPEHNT 589
 Qy 558 TGNMNC-NVATCTGEBYLVKOGIDISPMGLMKHVALACMIYVFLTIAYLKLLFLK 652
 Db 590 EBSRCHNVICTGEBFLMIQIGIDISLWSGFWENHVALVCTMIITLITVOLLQYK 645

RESULT 14

Q68HW7 PRELIMINARY; PRT; 646 AA.
 AC Q68HW7;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE ABC transporter ABCG3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RA Zhang W., Bissoco-Hagani S., Yam Y., Stanimirovic D.B.;
 RT "Cloning and characterization of ABCG3 in rat brain endothelial
 cells."
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF688113; AAT99308.1; -
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00005; ABC tran; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 SQ SEQUENCE 646 AA; 73170 MW; 3B5BE2D861F8DB2 CRC64;

Query Match 50.8%; Score 1703.5; DB 2; Length 646;
 Best Local Similarity 53.4%; Pred. No. 4.5e-102;
 Matches 350; Conservative 110; Mismatches 179; Indels 17; Gaps 8;

Qy 1 MSSNVEVPFIPVSGNTGFPATNSDLKAFTEGAVLSFHNIQYVKLKGFLPKRKYVE 60
 Db 1 MSSNDPVIPIPIERROGDLPELTNTSDPETTLTKAVLSFHNIQYVETQSSFP-PRQOTR 59
 Qy 61 -KEISNINIGIKKPGSLNIIIGPTGGKSSLDVLAAROPSGSLDGVINGAPRANFNC 119
 Db 60 VMERISISIGIMPGSLNIIIGPTGGKSSLDVLAAROPSGSLDGVINGAPRANFNC 119
 Qy 120 NSGVVQDDVWGLTVRENLOFSALRLATMTNHEKERNINRYIOELGIDKVDASKVG 179
 Db 120 TSGYIPQNDVVLGVTVADNLEFSALRLPMTVTDEKRRINEVLELHLEK----- 172
 Qy 180 TQFIRVSGGGERKRTSIGMELITDPSILFLDEPTGLDSTANAVLLIKMSKQGRIT 239
 Db 173 EONVPRKGRKRTSIGMELITDPSILFLDEPTGLDRTTDTVISILRRMSMGKTI 232
 Qy 240 FSIHQPRISIFKLPSTLLASGLMFGHQAQALGYESAGYHCEAVNMPDEFLLDIIN 299
 Db 233 FSIHQPRISIFKLPSTLLASGLMFGHQAQALGYESAGYHCEAVNMPDEFLLDIIN 292
 Qy 300 GDSSTVALNREED-SEATEIIPSPKQDKPLIEKLAIEYVNSFYETKALHQLSGEKK 358
 Db 293 GGFPE----DTEBDGHADENEFERQGVQYQKMIQMSPPYHATARADQLLGEQKL 348
 Qy 359 KKITVFKISYTSSTCHOLRWVSKSPXNLGNPQASIAQIIVTVVLGVIGAIYGLKN 418
 Db 349 DRSSAV-ETTCVTPFMHQIGMITRRSPKXNLGNPQASIAQIIVTVVLGVIGAIYGLKN 407
 Qy 419 DSTGIONAGVLFITTNOCFSSAVAVLFVVEKELFHEIYSGYRVSSTYGLGLD 478
 Db 408 VCTELQMRALILLGLGFCITSVTAGLEVIDQRFLEHRTSSTYGLGLD 467
 Qy 479 LPMTMLPSIIPTCIYVFMGLPKADAFVMMFTLMMVAASASALAIAGQSVASVT 538
 Db 468 IPRRLLPSTIIFLLIYFIAGLRTSVRGFTTFTIMLAISASSLSISGAGENVAAIT 527

Qy 539 LMTICFVEMMI FSGLVNLTIASMLWIOYFSIPRYGFTALQNEFLGONFCPLNAT 598
 Db 528 LVTIVYFVFMFLFSSLSLD-TGFLPVLWSIRKIFSIPHYGFRALLNNEFLGONFCPEYNT 586
 Qy 599 GNNPC-NVATCTGEBYLVKOGIDISPMGLMKHVALACMIYVFLTIAYLKLLFLK 653
 Db 587 EBSRCHNVICTGEBFLMIQIGIDISLWSGFWENHVALVCTMIITLITVOLLQYK 642

RESULT 15

Q68IH1 PRELIMINARY; PRT; 1039 AA.
 AC Q68IH1;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Similar to CA1379|CADP1 Candida albicans CADP1 ATP-dependent
 permease.
 GN ORFNames=DEHA0G113969;
 OS Debaryomyces hansenii CBS767.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycete.
 OX NCBI_TaxID=284592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;
 RG Genolevures;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neugebelle C., Talla E.,
 RA Goffard N., Francaul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barray S., Brancion S., Beckerich J.M., Beyne E., Bleykaten C.,
 RA Bollerame A., Boyer J., Catolico U., Confantolero F., de Darvar A.,
 RA Despons L., Fabre E., Fairhead C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicud J.M., Nikolaki M., Ozias S., Oziar-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Sweeney D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zenlin-Meyer M., Zivanovic I., Bolotin-Pukhara M., Thierly A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
 RT Wincker P., Souciet J.L.;
 RL "Genome evolution in yeasts."
 RL Nature 430:35-44(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS767;
 RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the ABC transporter family.

DR EMBL; CR382139; CAG90476.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:000524; F:ATP binding; IEA.
 DR GO; GO:0042626; F:ATPase activity; coupled to transmembrane m. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; F:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00005; ABC tran; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA_1.
 DR DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 DR DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR DR ATP-binding; EGF_2; 2.
 SQ SEQUENCE 1039 AA; 117062 MW; 51C861CAC3E092B4 CRC64;

Query Match 25.8%; Score 864; DB 2; Length 1039;
 Best Local Similarity 31.8%; Pred. No. 2.4e-47;
 Matches 219; Conservative 127; Mismatches 254; Indels 88; Gaps 13;

